7-2021

Accurate Simulation of Both Sensitivity and Variability for Amazonian Photosynthesis: Is It Too Much to Ask?

Sarah M. Gallup
Colorado State University, Fort Collins

Ian T. Baker
Colorado State University, Fort Collins

John Luke Gallup
Portland State University, jlgallup@pdx.edu

Natalia Restrepo-Coupe
University of Arizona

Katherine D. Haynes
Colorado State University, Fort Collins

See next page for additional authors

Citation Details

Gallup, S. M., Baker, I. T., Gallup, J. L., Restrepo-Coupe, N., Haynes, K. D., Geyer, N. M., & Denning, A. S. (2021). Accurate simulation of both sensitivity and variability for Amazonian photosynthesis: Is it too much to ask?. Journal of advances in modeling earth systems, 13(8), e2021MS002555.

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2021MS002555

Accurate Simulation of Both Sensitivity and Variability for Amazonian Photosynthesis: Is It Too Much to Ask?

Sarah M. Gallup[1], Ian T. Baker[2], John L. Gallup[3], Natalia Restrepo-Coupe[4,5], Katherine D. Haynes[2], Nicholas M. Geyer[2], and A. Scott Denning[1,2]

[1]Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO, USA, [2]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA, [3]Department of Economics, Portland State University, Portland, OR, USA, [4]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA, [5]School of Life Sciences, University of Technology Sydney, Ultimo, NSW, Australia

Key Points:

- Regression logic is reason to doubt the accurate climate sensitivity of predictions whose variability is realistic or higher
- A suite of models poorly reproduces eddy covariance estimates of Amazon rainforest gross primary productivity
- Highly seasonal models predict stronger primary productivity responsiveness to meteorology than is likely to be true

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. M. Gallup,
sgallup@colostate.edu

Citation:

Gallup, S. M., Baker, I. T., Gallup, J. L., Restrepo-Coupe, N., Haynes, K. D., Geyer, N. M., & Denning, A. S. (2021). Accurate simulation of both sensitivity and variability for Amazonian photosynthesis: Is it too much to ask? *Journal of Advances in Modeling Earth Systems*, *13*, e2021MS002555. https://doi.org/10.1029/2021MS002555

Received 1 APR 2021
Accepted 28 JUL 2021

Abstract Estimates of Amazon rainforest gross primary productivity (GPP) differ by a factor of 2 across a suite of three statistical and 18 process models. This wide spread contributes uncertainty to predictions of future climate. We compare the mean and variance of GPP from these models to that of GPP at six eddy covariance (EC) towers. Only one model's mean GPP across all sites falls within a 99% confidence interval for EC GPP, and only one model matches EC variance. The strength of model response to climate drivers is related to model ability to match the seasonal pattern of the EC GPP. Models with stronger seasonal swings in GPP have stronger responses to rain, light, and temperature than does EC GPP. The model to data comparison illustrates a trade-off inherent to deterministic models between accurate simulation of a mean (average) and accurate responsiveness to drivers. The trade-off exists because all deterministic models simplify processes and lack at least some consequential driver or interaction. If a model's sensitivities to included drivers and their interactions are accurate, then deterministically predicted outcomes have less variability than is realistic. If a GPP model has stronger responses to climate drivers than found in data, model predictions may match the observed variance and seasonal pattern but are likely to overpredict GPP response to climate change. High or realistic variability of model estimates relative to reference data indicate that the model is hypersensitive to one or more drivers.

Plain Language Summary Global climate models must accurately represent many processes, including the rate at which plants convert sunlight, water, and CO_2 into sugar. The enormous Amazon rainforest is extremely biologically productive, so the region strongly influences global CO_2 cycling. Measurements from instruments on towers in the Amazon, despite imperfections, seem to provide the most accurate estimates of rainforest plant productivity rates that exist. We compare tower estimates to 18 global models, focusing on subtle dry versus wet rainforest seasons. Generically, reality is more variable than models simulate. If a model accurately represents changes to an outcome when predictor variables change, predictions will have a narrower spread than is realistic. Accuracy errors can overwhelm the smoothing tendency, however. Models that add random uncertainty are an alternative in some simple situations, but the tendency toward narrowed predictions remains problematic for climate models. Modeled monthly plant productivity poorly matches tower estimates. About half the models have stronger seasonal cycles than the towers and half have less. By mathematical logic, the models with high seasonal spread must be overly responsive to inputs. As expected, the strongly seasonal models simulate greater change in productivity in response to variations in rain amounts than tower data suggest is realistic.

1. Introduction

Modeling the productivity of tropical forests is important for accurate climate predictions. The main reason is that the most productive biome in the world disproportionately drives global carbon and water cycles. Model estimates of current rainforest productivity vary by two-fold even after model differences in simulated precipitation are removed (Malhi et al., 2009; Poulter, Aragão, et al., 2010). The spread in simulations of current rainforest productivity is wider than for other biomes (Anav et al., 2015; Beer et al., 2010; Cavaleri et al., 2015; Friedlingstein et al., 2006; Jung et al., 2020; Mystakidis et al., 2016). Tropical forest decline has the potential to accelerate climate change (Allen et al., 2010; Levine et al., 2016; Malhi, Aragão, Galbraith,

et al., 2009). Any models that predict unrealistically weak responsiveness to weather imply unreasonable reassurance of continuing very high tropical productivity despite climate change. Excessive net weather responsiveness, on the other hand, will predict more change and likelihood of ecosystem collapse than is realistic.

As we will show (Section 3.4), in many earth system models (ESMs) tropical gross primary productivity (GPP) depends largely on four environmental drivers: temperature, precipitation, CO_2 concentrations, and top-of-canopy insolation. Each of these climate characteristics has been or soon will be altered in the Amazon. The simplest trends are for temperature. Rainforest temperature has already increased (Corlett, 2011; Gloor et al., 2015; Jiménez-Muñoz et al., 2013; Malhi and Wright, 2004) and will keep rising (Bathiany et al., 2018).

Rainfall in the Amazon is likely to become more variable (Chadwick et al., 2015; Feng et al., 2013). A tendency for dry periods to be more extreme is already observed (Fu et al., 2013; Gloor et al., 2013). Greater frequency of rainfall extremes reduces productivity because losses during droughts typically exceed gains during pluvials (Murray-Tortarolo et al., 2016; X. Wang et al., 2014), variability can reduce soil water recharge (Ryan, 2011) and floods can have opposing effects on productivity (Aragão et al., 2014; Brienen et al., 2016; Hawes & Peres, 2016). Studies reach differing conclusions about whether rainforests currently are net carbon sources or sinks either overall (Aragão et al., 2014; Brienen et al., 2015; Cleveland et al., 2015; Gatti et al., 2021) or specifically during drought years (Doughty et al., 2015; Feldpausch et al., 2016; Gatti et al., 2014). The direction of net change in Amazon rainfall remains uncertain (Gloor et al., 2012; W. Li et al., 2006, 2008; Poulter, Aragão, et al., 2010). El Niño-Southern Oscillation (ENSO) is a key control on tropical precipitation (Liu et al., 2017; Lopes et al., 2016; Panisset et al., 2018). A recent review concludes that so far models disagree about even a direction of change for ENSO frequency or amplitude (S. Yang et al., 2018).

Rising ambient CO_2 may be increasing rainforest GPP, but availability of water, nutrients, and suitable temperature may limit the extent of future change (Cox et al., 2013; Rammig et al., 2010). More CO_2 can increase productivity directly and reduce stress from water loss (Keenan et al., 2013; Swann et al., 2016), and can increase GPP by supporting higher leaf area (Z. Zhu et al., 2016) and extend the growing season (van der Sleen et al., 2015). Unlike in most of the world, in the tropics phosphorus rather than nitrogen is typically the most limiting nutrient (Castanho et al., 2013; Quesada et al., 2009; X. Yang et al., 2014).

Predictions about net change in GPP due to changing rainforest insolation are less clear even than for rain. Reduced cloudiness has slightly increased the light available within Amazon tree canopies (Barkhordarian et al., 2017), though whether the trend is either anthropogenic or robust is unclear (Wielicki et al., 2002). Pollution and tropical fires reduce incoming light. On the other hand, the aerosols scatter light deeper into canopies, increasing the productivity of subcanopy leaves (Butt et al., 2009; Rap et al., 2015; Yan et al., 2017). Fire's other simultaneous effects on GPP, including feedbacks that can increase post-fire surface temperature (E. A. Davidson et al., 2012), complicate sorting out how fire affects tropical GPP through light.

Data sets related to GPP, which are the foundation for assessing and improving model performance, have special limitations for the tropics. For remote sensing, tropical challenges are frequent cloud contamination, and global minima in revisit frequencies and portion of nadir views. Also, seasonal variability of light and temperature are stronger and therefore easier for satellites to measure outside of the tropics. In consequence, accuracy of satellite GPP estimates tends to be lowest for the tropics (Alemohammad et al., 2017; MacBean et al., 2018; Ryu et al., 2019). In situ data is therefore especially valuable for reducing uncertainty about tropical GPP (Cavaleri et al., 2015).

We explore the fidelity of modeled Amazonian GPP to eddy covariance (EC) flux tower data. We compare a suite of process and statistical models to six evergreen broadleaf forest (EBF) EC sites from the Large-scale Biosphere-Atmosphere Experiment in Amazonia. The first hypothesis is that each model accurately hindcasts the mean and variance of EC GPP both overall and at individual sites. Next, the sensitivity of GPP to current month's precipitation, temperature, and insolation for each model is summarized using linear regression. The degree to which this simple description fully characterizes the inner workings of each

model indicates the linear importance of weather to the models. We hypothesize that long-term site mean productivity, rain, temperature, and light drive most of the variation in simulated GPP.

This study features the inevitable trade-off for regression predictions between accurate variability and accurate sensitivity to predictors. By definition, models simplify, which means that any model omits some driver variable(s) and/or process complexities. When deterministic predictions are accurately sensitive to included predictors, unrealistically low variability is inevitable. Specific to GPP, the regression slopes for each weather driver reveal the model's effective sensitivity to climate change. We hypothesize the driver slopes will be reasonably accurate as indicated by similarity to EC slopes.

The tendency for predictions to be less variable than source data when responsiveness is accurate provides a tool to assess model sensitivity to drivers. Accurate or high outcome variability suggests excessive model response to changes in drivers. The greater the predictive power of weather for EC GPP, the greater the portion of dynamism that models heavily reliant on weather can accurately reproduce. From the logic, we hypothesize that models with high variability as indicated by seasonal amplitude greater than that of ECs are overly sensitive to weather.

Next, we assess the phase rather than amplitude of seasonality. We calculate the month of annual minimum productivity for each model as one descriptor of responsiveness to recurring drought stress. The null hypothesis is that EC timing matches that of modeled GPP. Alternatively, differences suggest potentially insufficient access to and/or presence of modeled soil moisture. Lagged and cumulative effects of rain are described qualitatively in the analysis section but data limitations preclude their inclusion in the numeric assessments. Finally, we illustrate more broadly the wide-ranging consequences for land surface models of the inevitable realism trade-off between variability of predicted outcomes and responsiveness to model drivers.

2. Methods

2.1. Model Data Sources

We compare EC GPP to 15 process models and three statistical models. The statistical models, Fluxcom, WeCann, and the vegetation process model (VPM), each have fared well in global accuracy intercomparisons. Multi-scale synthesis and Terrestrial Model Intercomparison Project (MsTMIP; Huntzinger et al., 2014; Wei et al., 2014) SG3 runs for 14 process models have common initial land cover maps, land use and land cover change, spin-up procedures, and atmospheric CO_2 and weather inputs. To MsTMIP we added SiB4 (Haynes, Baker, Denning, Stöckli, et al., 2019; Haynes, Baker, Denning, Wolf, et al., 2019), a recent major revision to the participating SiB3 model and which now has prognostic phenology. The evaluation period is limited to 2000–2010 to facilitate comparison of process models to statistical models that rely on satellite data. WeCann starts in 2007 with the earliest satellite data set for solar-induced fluorescence. WeCann is included only in basinwide comparisons because its shorter period for approximating seasonal cycles at tower sites means that the variability of WeCann GPP would appear artificially high relative to other models. Methods summarized in this section are described further in Text S2 (Andreae, 2002; Aragão et al., 2007; Bonan et al., 2011; Carswell et al., 2002; Chapin et al., 2006; Clark & Clark, 2011; Costa & Cohen, 2007; Domingues et al., 2005; R. A. Fisher et al., 2006; Goulden et al., 1996; Grace et al., 1995; Hutyra et al., 2007; Iwata et al., 2005; Khodaparast et al., 2008; Malavelle et al., 2019; Malhi, Aragão, Metcalfe, et al., 2009; Miller et al., 2004; Negrón Juárez et al., 2009; Oleson & Lawrence, 2013; Oliveira et al., 2005; Pyle et al., 2008; Rice et al., 2004; Saleska et al., 2003; Xu et al., 2015; Zhang et al., 2016).

To assess how well each model's behavior at EC sites represent its depiction of the whole Amazon, some basinwide comparisons are provided. The study area is northern South America. We refer to the world's largest rainforest as the Amazon although we do not use a strict watershed boundary. The study cells are limited to 42°–81°W and 12°N–21°S, excluding Central America both north of 7°N and west of 77.5°W. Selecting grid cells whose MsTMIP tiled plant functional types (PFTs) are at least 50% EBF limits the study to rainforest vegetation. Figure 1 shows the portion of each study cell that MsTMIP codes as EBF.

All the models in this study depict tropical photosynthesis in simplified ways. The models characterize each grid cell as either a single PFT or as one instance each of multiple cohabiting PFTs. The enzyme

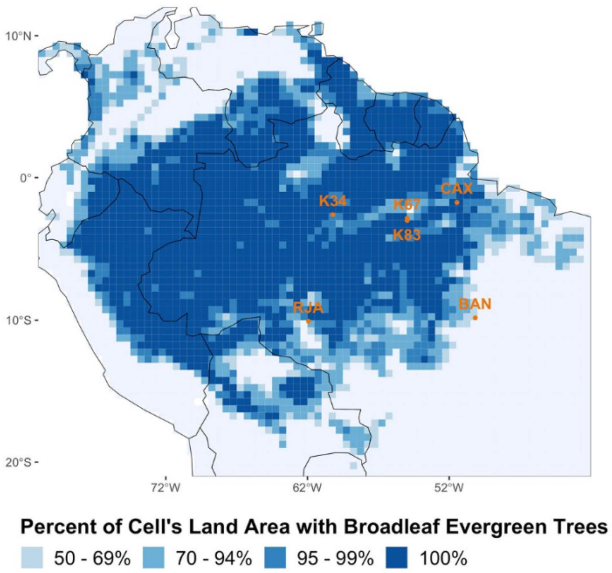

Figure 1. Study area, which includes all shaded cells. The Multi-scale synthesis and Terrestrial Model Intercomparison Project land plant functional types distribution in 86% of study cells is at least 90% evergreen broadleaf tree, or almost pure rainforest. Orange dots are locations of eddy covariance towers. Sites Tapajos Kilometer 67 (K67) and Tapajos Kilometer 83 (K83) are so close to each other that their location dots overlap.

kinetics models explicitly simulate only one or two representative leaves. The process models have no information about cell-level stand demographics, logging, herbivory, floods, or regional subsurface water flow. Also omitted are response to nutrient limitations and changes in photosynthetic efficiency due to leaf replacement that may vary seasonally and with drought. Some newer versions of these and other models do or soon may include some of the elaborations (J. B. Fisher et al., 2014; Keenan et al., 2012; Thomas et al., 2013). We assume nevertheless that ESMs neither will nor probably should ever represent every consequential influence on modeled natural processes. Instead, we ask what must be accommodated to the reality of missing predictors. And at the end of the study, we offer limited suggestions about how.

2.2. Eddy Covariance Data Source

The six EC sites to which modeled GPP is compared are Manaus Kilometer 34 (K34), Caxiuanã (CAX), Reserva Jaru (RJA), Rio Javaés-Bananal (BAN), Tapajos Kilometer 83 (K83), and Tapajos Kilometer 67 (K67) (Restrepo-Coupe et al., 2021). Sites are listed in order of average dry season length for 1998–2011, from 2.2 months at K34 to 5.2 months at K67. Annual precipitation as measured on site and averaged for the duration of each tower's operation ranges from 1,220 to 1,975 mm, in the same rank order as dry season length. Mean annual temperature is very similar across sites, ranging from 25.3 to 26.3°C. The dominant vegetation at each site is EBF. Figure 1 maps the tower locations. Table S1 gives additional site descriptions, and Text S2 includes a summary of some site particulars. Compared to the full range of modeled productivity across the Amazon basin, the GPP of EC cells is slightly higher than average (Text S3). The tendency is a methodological strength. Both satellites and EC data indicate that tropical forest GPP is the highest in the world (Anav et al., 2015). While ecosystems with lower productivity can be studied in multiple biomes, nowhere but in rainforests can the very top of the GPP scale be monitored in situ.

Some naming and unit conventions apply throughout the study. We define “site” as an EC location. “EC” refers more specifically to measurements made at a site and their derivatives. GPP estimates from each process or statistical model for the grid cell containing a site are matched to the months for which data exists at each tower. By “current weather” we mean light, temperature, and precipitation in the same month as GPP. Precipitation is monthly total in mm. Light is monthly mean top-of-canopy short-wave radiation under all sky conditions, in Wm^{-2}. Mean monthly temperature is measured in °C, and GPP in $gCm^{-2}d^{-1}$.

Unless noted, MsTMIP is the source of weather data used to assess modeled GPP responsiveness. Modeled GPP could instead be compared to on-site meteorology measured with each EC's instruments. Model simulations based on tower meteorology would provide cleaner comparisons than GPP calculated from weather streams that represent entire grid cells. However, the MsTMIP models used reanalysis data and there is not a parallel set of runs for specific EC cells forced with tower meteorology. Correlations of MsTMIP weather drivers for each month of EC operation with weather measured on site are 0.77, 0.63, and 0.46 for rain, temperature, and light, respectively.

EC towers are flawed benchmarks for GPP. Measured net ecosystem exchange is a small and imperfectly measured residual whose much larger offsetting components of GPP and ecosystem respiration must be modeled. A particularly thorny issue is lack of closure in energy budgets. Measured energy leaving a site does not equal measured energy entering (da Rocha et al., 2009; Jung et al., 2019; von Randow et al., 2004). Where GPP seasonal cycles are mild, as in the tropics, closure corrections introduce relatively more noise (Clark et al., 2017; Tramontana et al., 2016). These weaknesses are serious. Nevertheless, we take tower estimates to be the best reference data available. We accept their GPP responsiveness to individual drivers as true to the extent of being qualitatively strongly positive, strongly negative, or weak.

2.3. Statistical Methods Used

We compare statistical and process models to EC GPP mostly with simple statistics. Means and/or variances are contrasted, in some cases binned by values of a driver. While a few t-tests for the likelihood of true difference in population means are described, in most cases the divergences are so large relative to confidence intervals that they obviate formal testing.

Simulation models are ranked by the amplitude of their seasonal cycles. To describe annual GPP cycles across outliers for EC data that have missing data and uneven numbers of observation years per month, we fit Fourier series for each site and model. Earth's annual insolation cycle is sinusoidal, giving Fourier transformations inherent good fit for some ecological processes. A Fourier series recognizes sequencing, such as whether GPP peaks in the same month every year. Variance captures only the degree to which some individual monthly means are higher or lower than the overall mean regardless of time of year. Characterizing seasonal cycles of GPP with four pairs of Fourier terms is a compromise between overfitting and unrealistic simplification. The first pair of terms can be thought of as creating an annual cycle, the second allows for asymmetric shoulders, and the third and fourth provide for limited shaping of the annual peak and trough. We label the difference between maximum and minimum monthly GPP in a site's Fourier cycle as seasonal amplitude.

Regressions characterize each model's responsiveness to weather drivers. The equations allow the drivers' individual influences on GPP to be parsed and evaluated (Hamby, 1994). With differing future trends expected for each weather element, a model with retrospectively credible responsiveness to each driver is more likely to predict reliably. Equation 1 is an example of the form and describes EC GPP.

$$\begin{aligned} &\text{GPP} = -213 + \left(0.0043 * \text{Rain}\right) + \left(-0.012 * \text{Light}\right) + \left(0.74 * \text{Temperature}\right) \\ &P \text{ - values: Intercept, rain, and temperature} < 0.0001; \text{Light} = 0.11 \\ &\text{Adjusted } R^2 = 0.12; \text{Residual standard error} = 1.7;\ n = 260 \end{aligned} \tag{1}$$

Descriptive regressions allow comparisons across models that have greatly varying underlying forms and source data. Treating each month as an independent observation, the linear form is accurate in capturing a model's average responsiveness to weather during the period each EC operated. But if a model's sensitivity is non-linear, so that change in GPP depends not only on the magnitude of change but also the specific value of a driver, the linear form cannot fully describe model sensitivity to outlier values of weather nor, therefore, for a novel future climate.

The descriptive regressions do not reproduce the calculations by which each model derives GPP. For example, models based on enzyme kinetics have no internal regression coefficient for temperature, but instead multiple calculations involving temperature. In some models a key parameter is Q10 (Huntingford et al., 2013), the exponent for a rate multiplier to rubisco carboxylation per increase of 10°C. The value of Q10 may be derived from bench or field research. Q10 is not estimated directly for rainforests due to lack of source data, and because in theory the parameter is constant for all chlorophyll (but see Alster et al., 2020). Observationally based adjustments for each PFT's optimum temperature may occur. Higher temperature may affect photosynthesis rates by increasing vapor pressure deficit and surface evaporation rates, thus eventually acting also through soil moisture (Gloor et al., 2015). This partial list of mechanisms by which temperature can affect GPP illustrates that nudging a process model for some desired change in net sensitivity evokes tugging on the universe.

2.4. Deterministic Predictions Have Less Variability Than Source Data

A model can predict only what it "knows" about. Model predictions simulate the subset of the source data's variability that the included inputs and selected mathematical forms can explain (Vogel, 1999). Models cannot simulate the portion of actual dynamism that omitted influences determine (Farmer & Vogel, 2016). Model simplifications reduce the variance of predicted outcomes relative to the variance of measured outcomes, a characteristic we label as "flattening." Text S1 (R. Davidson & MacKinnon, 2004; Greene, 2012) formally expresses the statistical origin of flattening.

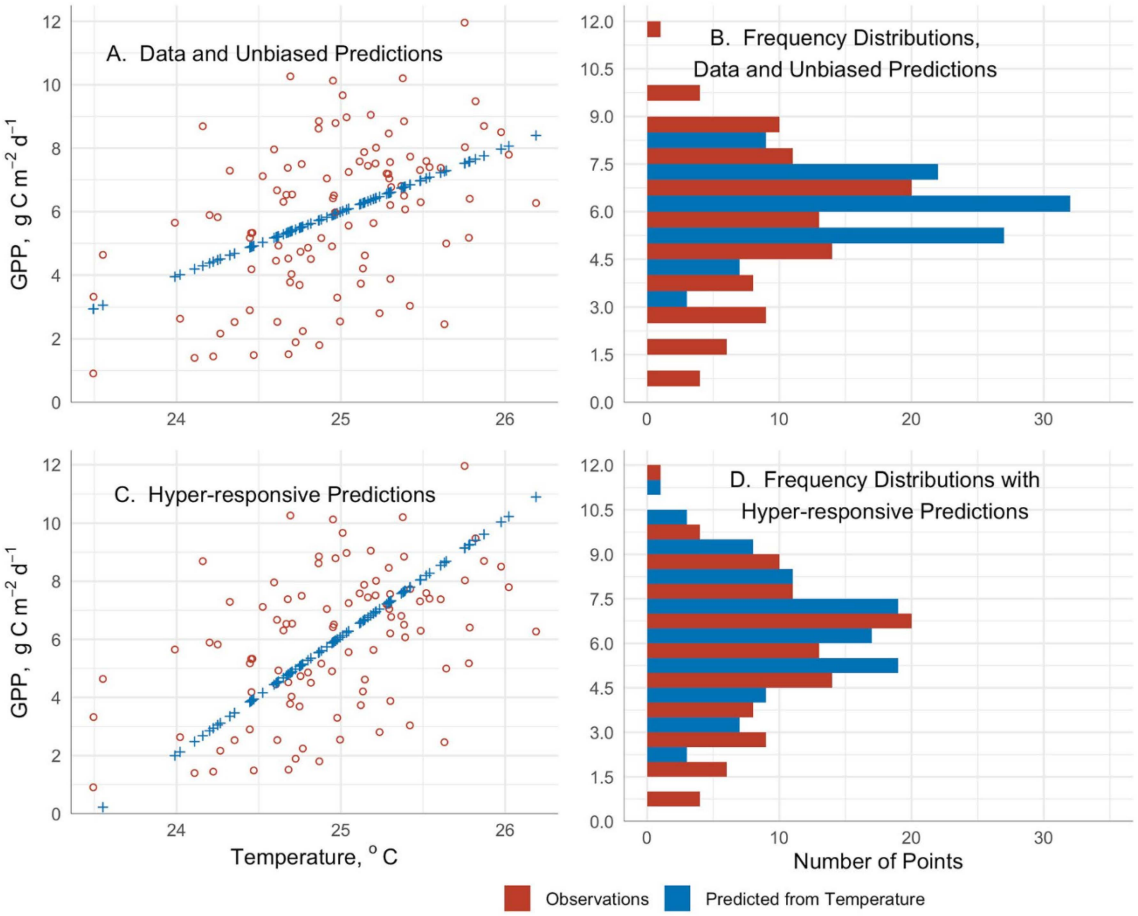

Figure 2. A random sample of simulated observations illustrates that the dynamism of derived regression predictions is necessarily less than the dynamism of the source data. Panel A shows the linear effect of temperature on gross primary productivity (GPP) for the underlying data and in the regression predictions. The frequency distributions of observed and predicted GPP from the unbiased regression in panel A is summarized in panel B. Panels C and D show predictions from the same source data for the same temperatures and with the same predicted mean GPP but with an arbitrarily doubled regression slope. While simulated variance in panels C and D better reproduces the source data's variance, the cost is predictions that overstate the consequences for GPP of a unit change in temperature.

An example with randomly generated data illustrates flattening. If GPP for some location comes with no additional information, the best estimate of any other GPP value is the observed average. The GPP observations are not identical and have a non-zero variance, yet the best predictions are identical with zero variance. If matched temperature data exist and the relationship is approximately linear, a regression gives improved predictions (Figure 2a). But as the contrast in spread of blue versus red frequency bars in Panel B displays, predicted GPP is still less variable than the measurements. Omitted influences underlie the residual errors.

The label "variance of predictions" may mean either of two concepts. For the variance of prediction coefficients, often shown as a confidence interval, smaller is preferable. This study addresses instead the spread across a model's predictions. More realistic is better, which often means bigger. The generic concept of simulating a higher proportion of natural variability we refer to as model dynamism. Simulations of responses to seasons and weather create some but not all of the dynamism in GPP models.

There are two ways that a model with a fixed form and set of predictors can generate predictions with more dynamism than does an accurate regression of the source data. First is deliberate addition of random variation. Stochastic modeling is considered in the discussion. Otherwise, this study addresses only deterministic models, in which a set of predictor values generates the same outcome each time the model runs. The second-way model predictions can have higher variability than source data is inflated driver responsiveness. Sensitivity, or responsiveness, is marginal change in outcome per unit change in a predictor (Friedlingstein

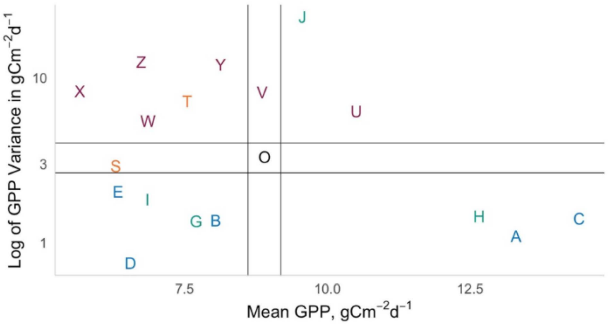

Figure 3. The mean values and variability of gross primary productivity (GPP) predicted by 17 models for 1° cells that encompass an Amazon flux tower show large differences from eddy covariance (EC) benchmark values. The 17 model values are designated by letters, with names listed in Figure 4. The parallel vertical and horizontal lines are bootstrapped 99th percentile confidence intervals around flux tower GPP metrics ("O"). All the 17 models have either means or variances outside the intervals, and all but models "S" and "V" have values falling outside both confidence intervals. The confidence bounds reflect only the distribution of calculated monthly means and do not include uncertainty in EC GPP. The y-axis has a $\log_{10}$ scale.

et al., 2006; Hamby, 1994). In a regression, responsiveness corresponds to slope coefficients.

Excessive responsiveness can more than offset the underlying influence of flattening. Figure 2 illustrates the mechanism. Panel (a)'s predictions represent true responsiveness given the source data. GPP responsiveness to temperature has been artificially doubled in panels (c) (individual data points) and (d) (frequency distributions). The result of the introduced error is wider spread in GPP predictions along the *y*-axis in panels (c) and (d). Because the slope in panels (c) and (d) is inaccurate, the consequences for simulated GPP of the warmest temperatures are excessive and vice versa. Exaggerating a true slope by a factor with absolute value greater than one increases dynamism by the square of the constant (Farmer & Vogel, 2016), or counteracts flattening. A multiplier with absolute value smaller than one underestimates responsiveness.

Because all models simplify, flattening is a characteristic of all deterministic numeric predictions. Flattening occurs with non-linear equations, transformed drivers (Text S1) and, like entire ESMs, in complex combinations of equations with feedbacks. Flattening also can take the form of hard-coded parameters that derive from a regression fitted to observations and are "essentially a smaller model within the larger model" (Dahan, 2010). If ESM sensitivities to drivers are accurate, then predicted outcomes will be less dynamic than actual outcomes.

Flattening can be a diagnostic tool. If a model's predicted variance exceeds the variance of accurate and representative benchmark data, then offsetting excessive sensitivity to one or more drivers exists. Because by definition models simplify, even accurate variability of outcomes indicates excessive dynamism. A side effect of the internal interconnections in ESMs is that the on-going process of improvements and fine-tuning involves complex trade-offs (Koven et al., 2013; Mauritsen et al., 2012). The magnitude of ESM modeling accomplishments to date and the difficulty of making improvements reinforce the usefulness of clear, simple diagnostics like flattening.

3. Results

3.1. Modeled GPP Means and Variances Differ Grossly From EC Estimates

The optimistic hypothesis that simulated GPP mean and variance match EC estimates is easily rejected. The models differ strikingly about both how productive the rainforest is overall, and how much its GPP varies from month to month (Figure 3). Averaged across the six sites and the months of each tower's operation, nearly all model estimates for both mean GPP and variance are well outside the lines in Figure 3 that mark 99th percentile confidence limits of EC GPP. The predictions from every model have a mean, variability, or in most cases both, that are inconsistent with the flux tower data. Viewed in greater detail as correlations at each site (Figure S7), the fit between individual study models and EC GPP ranges from −0.16 to 0.45 with a grand mean of only 0.12. Squaring the mean correlation for the model with the closest fit indicates that no model explains more than 15% of EC GPP variability. Mean correlations with EC GPP for four models are statistically indistinguishable from zero.

Site-level GPP (Figure S8) shows similarly low connection between study models and ECs. For individual sites, an average of only one or two models are credible matches to mean EC GPP (Figure S9). At least one model severely underestimates mean GPP at each site while the mean for at least one other model is more than twice EC GPP. For the most part which models are outliers differs across sites. However, two process models' (Models A and C) simulate higher mean GPP than ECs for every site, and one model (X) has consistently lower means. No model simulates GPP variance for every site that is within wide EC confidence intervals. Variance for one statistical model is credible for five of six sites. Site-level mean correlations of study models to tower GPP ranges from −0.32 to 0.66 (Figure S7). CAX and K67 have especially weak

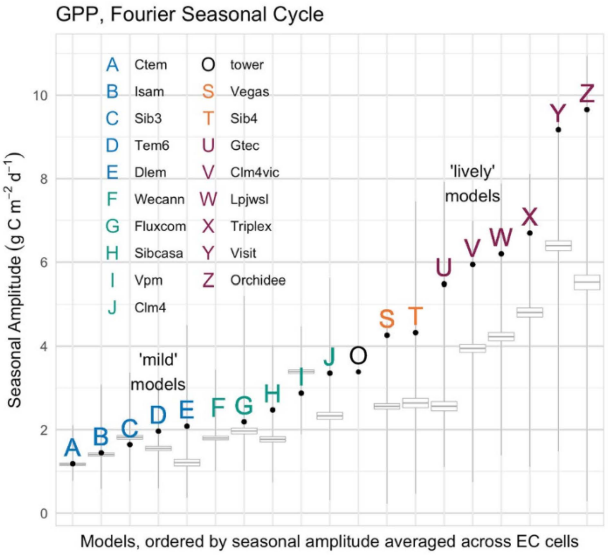

Figure 4. Models differ widely in the strength of their simulated gross primary productivity (GPP) seasonal cycle both at six eddy covariance (EC) sites, whose average cycle amplitude is shown with black dots, and across the Amazon, whose distributions of cell-level seasonal amplitudes are shown in gray. Seasonal amplitude is the annual maximum minus minimum of a four-term Fourier transform characterization of mean monthly GPP. Models are ranked on the x-axis by mean seasonal amplitude at tower sites, with EC seasonal amplitude falling squarely in the middle of the range. Gray boxes show a 95% confidence interval around the mean seasonal amplitude for all Amazon rainforest cells in all months. Whiskers on the gray boxes represent the tenth and ninetieth percentiles of cell amplitudes. Model F has no EC site mean, as explained in the Methods section, and its rank is an approximation. Colors of letters indicate how a model's amplitude compares to the EC amplitude, both averaged across all six sites. The nine "mild" models with weaker mean seasonal cycles than ECs are shown in blue or green. The eight "lively" models whose cycles are stronger are colored red or orange. The intensity distinctions for blue and red break at one standard deviation from the EC mean.

matches, with negative correlations for 12 and 14 of the 17 models respectively. The most closely simulated site is RJA, where EC GPP is especially dynamic and average correlation across all models is 0.66.

A mathematical artifact does not explain the relatively high variance of some of the study models. All else being equal, data sets with larger magnitudes have larger variance. The tendency would suggest that the variance of models with low mean GPP would be lower than for models with high mean GPP. But half the models in Figure 3 show the opposite pattern, pairing higher variance with lower mean GPP than the EC towers or the reverse. The opposing tendencies mean that for these GPP models, whatever causes differences in means does not explain dynamism. Instead, the causes of differences in responsiveness of model GPP to drivers needs to be considered directly. Before an exploration of model responsiveness, the next section assigns a descriptor of model dynamism that is more robust than variance.

3.2. Seasonal Cycle Amplitude Characterizes a Model's GPP Dynamism

Models can be ranked by the magnitude of their Fourier seasonal amplitude averaged across the six sites. Most notable is how widely the seasonal swings differ across models, by a factor of 8.2 (Figure 4). The difference means roughly that trees as simulated in Model Z vary in productivity eight times as much during a year as do the same trees as simulated in Model A. EC seasonal amplitude, 3.4 gCm^{-2}d^{-1}, is near the middle of the range. GPP for the mildest model varies during the year a third (0.35) as much as does EC GPP. Mean site amplitude for the most strongly responsive model is almost triple (2.9 times) that of the ECs.

Only one model (J) would switch between the categories of mild versus lively if rankings were determined by variance instead of seasonal amplitude. The model's very high overall variance (Figure 3) is due to its puzzling GPP close to zero for several grid cells near K67 even though its GPP is well above average at most other sites. When variability is measured instead as mean seasonal amplitude, Model J has the closest match to tower GPP (Figure 4).

Mean amplitudes have large uncertainty because each summarizes only six data points. Based on a t-test, only the liveliest model's amplitude is outside a 95% confidence interval around the EC mean. The K34 and K67 sites are near enough to each other to have somewhat similar climate. Calculating the standard deviation of mean EC GPP with four degrees of freedom rather than five to reflect possible pseudoreplication, no model is outside the credible interval. The sample size precludes tightening the confidence interval by bootstrapping. Consistent with differences in sample size, a model's mean seasonal variability for six EC sites is on average 1.2 times larger than its mean seasonal variability averaged across hundreds of study cells.

Strong disagreement about the strength of the rainforest seasonal cycle is evident at the finer spatial scale of individual EC sites, and at the wider scale of the entire Amazon rainforest. At no individual site is a very mild model's mean amplitude larger than 4 gCm^{-2}d^{-1} (Figure S10, blue letters). In contrast, few of the most responsive models (Figure S10, red letters) have a seasonal amplitude below 4 gCm^{-2}d^{-1} at any site. Tendencies across the entire basin also show marked differences in the magnitude of the seasonal cycle. ECs' differing periods of operation preclude a temporally matched comparison between a model's basinwide and site tendencies, so cell means for the full study period are compared. Average seasonal amplitude for all 1° study cells ranges from 1.2 to 6.4 gCm^{-2}d^{-1} for the most to least seasonal models, respectively (Figure 4), which creates differences in mean simulated seasonality of half an order of magnitude. The extent to which the EC cells are typical of the modeled Amazon as a whole decreases with seasonal amplitude. For the mildest

models, shown in blue in Figure 4, EC amplitudes are roughly representative of the entire basin. For all the strongly lively red models but one, EC sites have moderately stronger seasonality than do basinwide means.

Seasonal cycles are critical to the overall variability of simulated rainforest GPP. Interannual amplitude calculated as the difference between highest and lowest year's mean GPP from 2000 to 2010 for individual models ranges from 0.1 to 1.2 $gCm^{-2}d^{-1}$ (Figure S11). Models' mean cell-level seasonal amplitudes across the basin are several times larger, ranging from 1.3 to 6.0 $gCm^{-2}d^{-1}$. The similar range across models in Fourier seasonal cycle amplitudes for only EC cells, 1.2 to 9.7 $gCm^{-2}d^{-1}$, in contrast, approximately equals the grand mean of monthly GPP across all models, 8.9 $gCm^{-2}d^{-1}$. For GPP in the Amazon, determinants of change within a year explain much more about a model's tendencies than do determinants of its interannual variability.

3.3. EC GPP Barely Responds to Current Weather

Our hypothesis is that a simple linear combination of rain, temperature, and light largely describes monthly mean GPP. If so, annual weather cycles that exist even in tropical rainforests (Girardin et al., 2016) might logically also determine simulated annual cycles of GPP. Site-level differences account for most of the spread in the EC data set. Disaggregating the effects of individual drivers (Chevan & Sutherland, 1991) shows that site-level differences account for 81% of Equation 1's predictive power. Removing the influence of site effects leaves only 19% of the predictive power, or just one-eighth (12%) of the total dynamism in EC GPP, explained by the current month's weather. Contrary to our hypothesis, current weather has little influence on EC GPP.

At the limit, every characteristic of a site is either the current or the lagged effect of some processes (Bloom et al., 2020). For GPP, deep legacy effects include topography, a site's geologic parent material, and adaptations of resident species. A key biotic difference between sites is species assemblage in the spectacularly diverse tropics. An example of an abiotic influence that may affect site means is local and regional geology, which over millennia affects soil depth and capacity to store water, soil fertility, flooding, subsurface hydrology and more. Some of these influences are at present impossible to parameterize well across the entire globe, and for an ESM constitute impenetrable statistical noise.

An alternative regression form can isolate the unspecified reasons for different site means. Individual intercepts, also called fixed effects, represent site differences in the outcome that are both constant over the duration of the data and not explained by included drivers. Computationally, site effects for GPP represent the differences between each site's mean GPP and the overall mean. Site means summarize the lagged effects of all the processes that operate too slowly to cause dynamism at the time scale of the source data, in this study approximately a decade. The model ascribes its remaining simulated dynamism to faster processes, of which current weather is a subset. Including separate intercepts in the regression allows a focus on responses to included drivers that are independent of long-term site differences. Fixed effects have the cost of overlooking how accurately each site's mean GPP, or baseline productivity, may be modeled. Equation 2 adds site intercepts to Equation 1.

$$\begin{aligned} &\text{GPP} = \text{Six site intercepts} + \left(0.0054 * \text{Rain}\right) + \left(0.019 * \text{Light}\right) + \left(0.52 * \text{Temperature}\right) \\ &P\text{ - values: Rain and temperature} = 0.00; \text{Light} = 0.01 \\ &\text{Adjusted } R^2 = 0.59; \text{Residual standard error} = 1.2; n = 260 \end{aligned} \tag{2}$$

The Akaike Information Criterion (AIC; Akaike, 1974) balances an equation's ability to explain source data against an offsetting penalty for a large numbers of predictors. AIC provides a numeric method to judge the power of individual regressors. Based on AIC, all three drivers in Equation 2 should be retained. On average, more rain, more light, and higher temperature all increase EC GPP. This differs from Equation 1, where light's negative coefficient was not significant. Compared to Equation 1 with a single intercept, the fixed effects form explains almost five times as much of the overall dynamism of GPP.

Figure 5 illustrates flattening in GPP predictions. The graph compares EC GPP on the *x*-axis to predictions from Equation 2 on the *y*-axis. Both overall and for individual sites, the spread of point values is narrower on the *y*-axis than on the *x*-axis. The raw r^2 for Equation 2 of 0.61 unadjusted for degrees of freedom (adjusted = 0.59) indicates that the regression terms including site intercepts determine 61% of EC GPP dynamism.

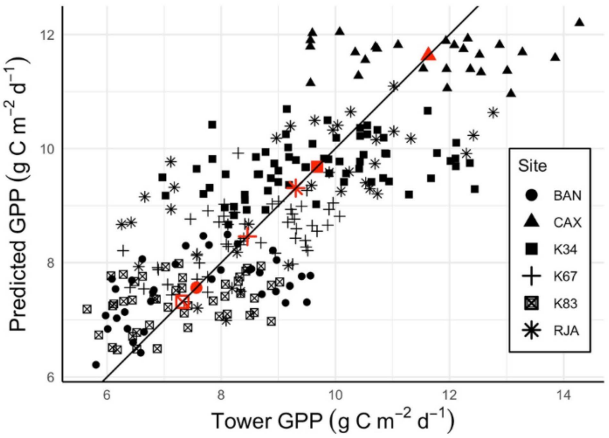

Figure 5. Monthly eddy covariance (EC) gross primary productivity (GPP) compared to values predicted from a regression of EC GPP on Multi-scale synthesis and Terrestrial Model Intercomparison Project rain, light, and temperature using site intercepts. Site fixed effects show that differences between sites are the main source of the prediction's power. Individual intercepts force each site's mean predicted value to equal the mean EC value (red symbols), while weather drivers determine a prediction's distance from the site mean. Among the points for any specific site, indicated by a unique symbol, there is minimal clustering around the 1:1 line of perfect fit.

The regression equation does not have information about the undefined group of influences that explain the remaining ~40% of observed differences in monthly GPP. So while the underlying EC GPP data points have a variance of 3.3 gCm^{-2}d^{-1}, the variance of the regression predictions is only 61% as large, or 2.0 gCm^{-2}d^{-1}. Graphically the contrast is apparent as vertically "flattened" point clouds.

If EC meteorology is used rather than MsTMIP weather, the regression fit with site effects degrades slightly (r^2 = 0.54). Different from Equation 2 that uses MsTMIP weather, the only terms whose p-value is ≤0.10 are four site intercepts and the slope for rain. GPP is statistically unrelated to either light or temperature. Why site-specific weather should be less predictive than regional weather is unclear. Perhaps soil moisture is influential, reflects regional recharge, and overwhelms highly localized rainfall differences. A geographically broader weather summary might more accurately represents conditions across ECs' full footprints than does weather at point locations chosen to represent the upwind area (but see Chu et al., 2021). The unexpectedly weaker fit with site weather is convenient, however. Errors in representing the true values of the drivers cause attenuation bias in regression coefficients, or weaker sensitivity. If MsTMIP weather were a worse fit than site weather, comparisons of EC driver sensitivities to model sensitivities would be less straightforward because there would be no simple way to know how much of the loss of fit were due to local weather being more representative of the site.

3.4. Lively Models Respond More Strongly to Current Weather Than Mild Models

The vigor with which some models respond to rain, light, and temperature contrasts sharply with EC GPP's subdued shifts. The regressions whose responsiveness slopes appear in Figure 6 parallel Equation 2, but use z-score transformations of predictors so that each slope represents change in GPP in response to a change in the driver of one standard deviation. Across all the study models, statistically significant coefficients (Figure 6a) range from −0.26 to 2.2 for rain, for temperature (panel (b)) from −1.8 to 0.26, and for light (panel (c)) from −0.81 to 0.69. EC responsiveness to each driver measured as change in absolute GPP is 0.6, 0.5, and 0.3 gCm^{-2}d^{-1}, respectively, while outlier models simulate much larger changes. Ambient CO_2 is an insignificant predictor of historical site GPP in the EC data and in nearly all models (Figure S12) and is not included in Figure 6's regressions.

The data generally support the hypothesis that a simple linear regression characterizes the inner workings of each model, though not for ECs. Current weather plus site means determine more than half of the change in GPP at a particular location for all but two models. For the mild models as a group, mean r^2 is 0.69 and residual standard error (RSE) averages 8% of site GPP (Figure S13). For the lively models, mean r^2 is 0.58 and average RSE is 24%. The simple equations describe the inner workings of seasonally mild models even more fully than they do for lively models.

Among the GPP drivers, rain is the most consistent predictor across models. GPP increases with rain in all but two models (Figure 6a), with one of the negative responses not statistically significant (p = 0.08). The magnitude of rain responsiveness ranges widely, however, from −0.3 to 2.2 gCm^{-2}d^{-1} of GPP per 120 mm increase in a month's precipitation. Model ranks for rain slopes almost match the order of seasonal amplitudes. EC responsiveness to rain sits solidly in the middle. For rain, the main difference among models is the response strength.

All three of the current weather elements substantially affect GPP. Absolute slopes for each weather element averaged across all models differ by less than a factor of two: 0.75 for rain, 0.46 for temperature, and 0.40 for light. For temperature and light, models disagree about both the direction and strength of GPP responses. Two patterns emerge in model responses to brighter months. First, fully a third of slopes are statistically inseparable from zero, meaning that on average light has no consistent directional effect on a

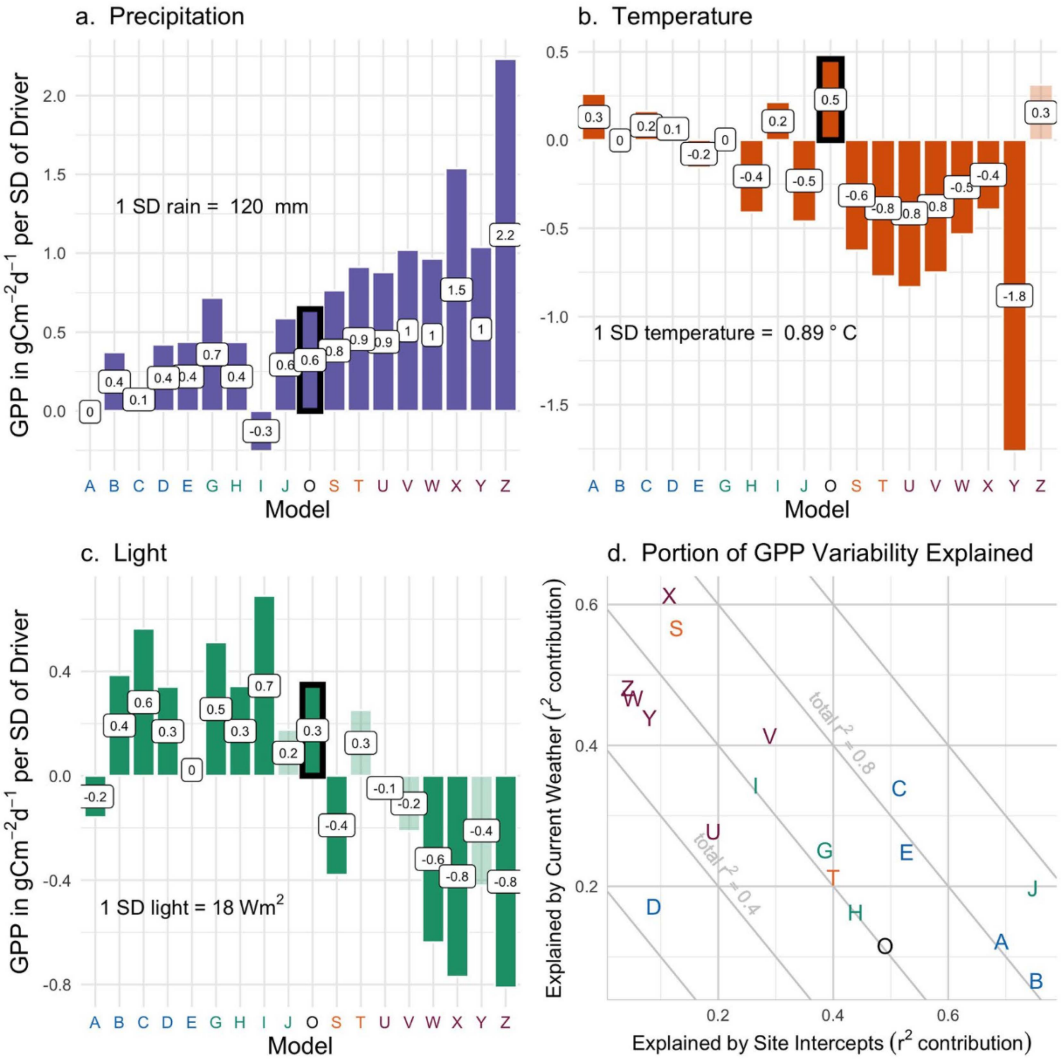

Figure 6. Among the study models, gross primary productivity (GPP) responsiveness to current month's weather at six eddy covariance (EC) sites as characterized by descriptive linear regressions differ widely for both individual weather elements and the magnitude of current weather's contribution to predictions. Regression slopes for precipitation (panel (a)), temperature (b), and radiation (c) are in units of standard deviation across all sites and months. For example, a regression slope of +0.5 indicates a 0.5 gCm^{-2}d^{-1} increase in GPP for an increase of one standard deviation in the driver. Pale bars are coefficients whose *p*-value exceeds 0.05. EC responsiveness is outlined in black. Panel (d) shows for each model the contribution to total r^2 from site intercepts versus current weather. For example, for ECs, Model O, site intercepts, which are displayed on the *x*-axis, explain 49% of variation in GPP while rain, temperature, and light together, shown on the *y*-axis, explain 12%. The point in panel (d) for Model O, whose descriptive regression explains a total of 61% of GPP variation, falls close to the gray line that marks total model r^2 of 0.60. Portions of variability explained are sums for site intercepts and for current weather respectively of change in total r^2 due to adding individual predictors, averaged across all permutations of sequencing for additions.

model's GPP. Second, most mild models simulate GPP increases with more light, while most lively models simulate GPP declines. EC GPP shows a moderate positive response to light. For temperature, models' mean GPP change ranges from −1.8 to 0.5 gCm^{-2}d^{-1} per approximately 1°C of increased warmth. Temperature slopes for four models are not statistically significant. Mild models tend to respond weakly to temperature increases, while in lively models on average GPP declines. In strong contrast, EC data suggest that within historically observed ranges, on average rainforests have thrived with more heat. For anticipating the effects of climate change on rainforests, the question of how rising temperatures will affect plant productivity is critical. We will return to the topic in the descriptions below of related research and of non-linearity in model responses to weather.

The principle of flattening indicates that strongly seasonal models, with higher variance than the benchmark EC data, on average respond excessively to weather. Figure 6 confirms that more lively models tend to have higher slopes than ECs for each weather driver. Lively models anticipate that a changing climate will have stronger effects on plants than EC data suggest is realistic.

Overall, one of the most striking differences among the models is the time scales that most affect GPP for a given location. The relative weight that each model ascribes to site differences, or to very slow processes, equals the percent of variance that site intercepts explain. Models close to the *y*-axis in Figure 6, panel (d), imply a view that most of the dynamism in tropical GPP is due to processes whose consequences vary during the eleven-year study period. For ECs, site means explain half (49%) of the overall differences in GPP values, or four times as much as does current weather. For mild models, site intercepts explain more of GPP's variance (mean = 49%) than does weather (mean = 21%). For lively models the opposite is true; Site intercepts explain a smaller share (mean = 16%) than does weather (mean = 43%). The explanatory strength of site influences as a percent of the influence of monthly weather variations differs for the most extreme models by more than two orders of magnitude, from 8% for Model Z to 1,136% for Model B. Mild models suggest that knowing a site's long-term fixed characteristics is relatively more important for explaining GPP than are influences that vary month-to-month, while lively models imply the opposite.

A model's responses to drivers in the six cells with EC towers are generally similar to its responses across the Amazon (Figures S12 and S13). The likeness suggests that assessing model responses for the Amazon by comparing them with EC estimates of GPP is a reasonable application of scarce benchmarking data. But the similar means do smooth across considerable spatial differences (Figure S14). The most striking spatial pattern is that for most models a simple regression describes change in GPP almost completely in some areas yet explains little in others.

3.5. Models' Differing Rainforest GPP Non-Linearities Are Not Benchmarked

Generically, regressions that describe model output lack some of the sources of uncertainties that can muddy regressions that describe field observations. For a model, all of the outcome values for the study domain typically are available, yielding a census with no sampling errors nor errors in measuring outputs. Some or all of the inputs to model calculations may be known exactly, as is MsTMIP weather. Only two sources of stochastic noise remain in the regressions that describe modeled GPP: omitted drivers and misspecification. While error due to omitted drivers embodies real-world dynamism, measurement error, sampling error, and misspecification of mathematical form represent the data and model's uncertainty about included aspects of the real world (Vicari et al., 2007). This section considers alternative specifications.

Interaction terms in a regression indicate interdependent consequences of simultaneous changes in multiple drivers. For example, lack of rainfall may impede GPP more if temperatures are simultaneously warmer. Adding interaction terms to the benchmark regression that describes EC GPP minimally increases its total explanatory power, while diluting evidence about weather's influence. The same predictors as Equation 2 were used plus all pairwise interaction terms for the three weather drivers: rain times temperature, etc. The resulting regression (not shown) has 5% more explanatory power than the model without interactions ($r^2 = 0.62$). But no single or interaction weather term has a significant slope. The models generally embody the same understanding, with increases in explanatory power due to weather interaction terms that range from 0% to 15% and have a mean of only 4%.

Shifting focus from interaction terms to non-linear forms for individual drivers, uncertainty about GPP response to specific temperature ranges as indicated by differences between models is greatest for the warmest observed temperatures (Figure 7). For most models, GPP differences at low temperatures compared to median temperatures is limited, very roughly a third of a standard deviation as shown on the *y*-axis. GPP at cells' high temperatures during the study period, on the other hand, deviate strongly. For anticipating the future, ESM responses to high temperatures are particularly important (Cavaleri et al., 2015). GPP rises continuously with temperature in about a third of mild models. The remaining models eventually alter from rising to declining plant productivity with increasing temperature, implying that a threshold exists above which rainforest productivity suffers in high heat. Several of the mild models simulate that rainforest is at its most productive when the mean temperature for the month is highest, while lively models depict

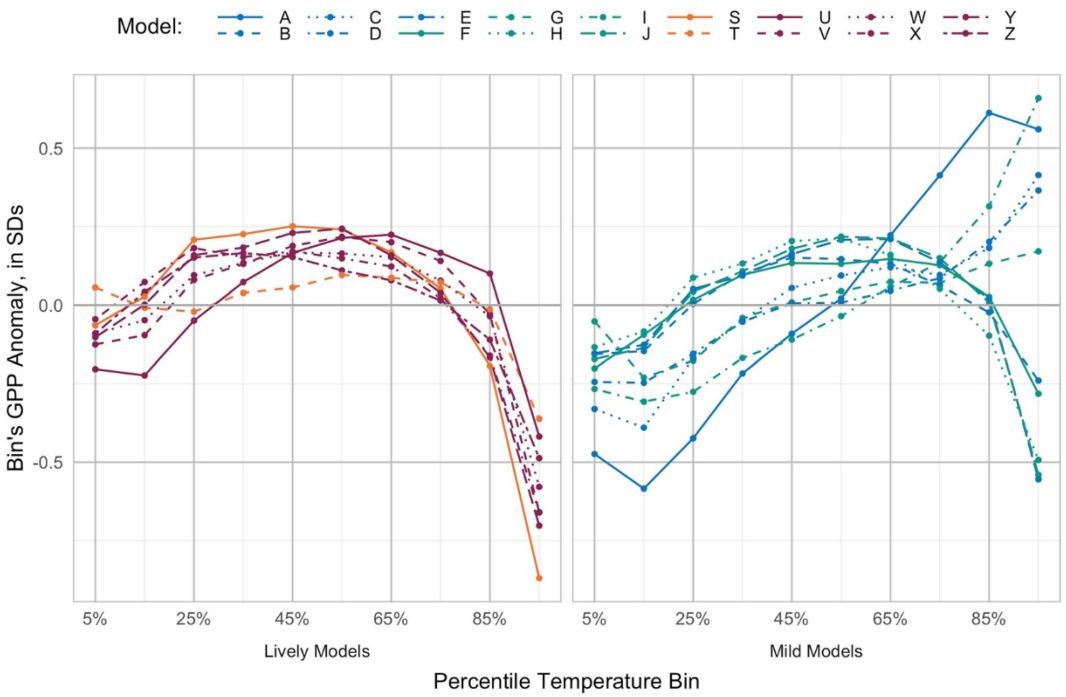

Figure 7. Non-linearity in modeled cell-level gross primary productivity (GPP) responses to temperature across the Amazon basin show that at high temperature, GPP falls markedly in lively models while the response of mild models has little commonality. Lively models are in the left panel and mild models in the right. On the *x*-axes, monthly mean temperatures are grouped by deciles basinwide. GPP on the *y*-axis is displayed as z-scores to remove model differences in mean and amplitude of variation for each cell. What remains is the degree to which a model's response to extremes of temperature is anomalous compared to the model's responses to currently more typical temperatures.

a starkly opposite conclusion. Mild Models H and J resemble the lively models in this respect, simulating their very lowest GPP at peak temperatures. It is an important and unresolved question whether tropical plant productivity will rise or fall as high temperatures become increasingly common.

Like responses to temperature, model responses to rain and light also are non-linear (Text S4; Dahlin et al., 2017; Goulden et al., 2004; Rogers et al., 2017). Models with the strongest, steepest GPP response to increasing rain tend to have only modest response to increasing light and vice versa. Most of the lively models respond strongly to rain but have below-average GPP in the brightest months. In contrast, most mild models simulate their highest rainforest GPP with typical, middle decile, rain amounts, and below-average GPP in the wettest months. The tendencies imply that according to the lively models the rainforest tends to be a bit thirsty, while the mild models reflect a premise that thus far, climatology is approximately optimal.

Resolving a non-linear regression form for assessing the accuracy of a model's optimum values of each driver would require more months of EC data. It is possible that EC GPP at finer temporal resolution would better reveal interactive and/or non-linear responses. First, the numbers of daily or hourly means are far larger than the number of monthly values, providing more degrees of statistical freedom. Second, outlier weather, which is especially important for defining the shape of a non-linear response curve, is more likely to be represented in daily data. However, MsTMIP's monthly resolution precludes direct comparisons.

3.6. Lively Models Simulate Strong, Rapid Drops in Dry Season GPP

Comparing the annual cycle of relative GPP to the dry season roughly indicates how quickly a model suggests that depletion of stored water stresses plants. It is possible for the modeled timing of the season when a forest experiences its greatest stress to be correct in most of the world but have major inaccuracies in the very wettest regions (Collier et al., 2018, Figure 5d). Accuracy in simulating the month of lowest GPP is uneven for all models. Across the six sites, a model's month with lowest GPP is on average 2.6 months different

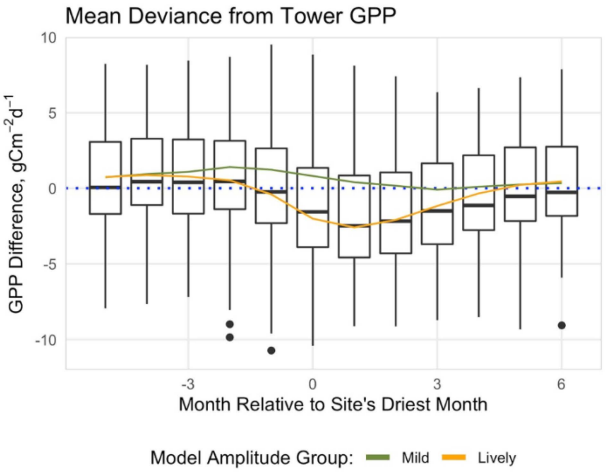

Figure 8. Seasonal deviations of gross primary productivity (GPP) from eddy covariance (EC) estimates for models grouped by seasonal amplitude show that model divergences from EC GPP are especially large for lively models in the latter half of the dry season. On the *y*-axis, zero represents an exact match to EC GPP. Green and orange lines are mean deviances for mild and lively models respectively. Zero on the *x*-axis is the calendar month at each EC site with lowest average rainfall, with one month before the driest month shown as −1, two months after as +2, etc. Month crossbars show median tendencies of all models as a group compared to ECs, with boxes enclosing the 25th and 75th percentiles of all models' GPP deviances from the EC estimates. Dots indicate outlier models.

from the EC's minimum month (Figure S3). Every model matches the time of minimum EC GPP to within one month for at least one site. But with one exception, every model also simulates the timing of minimum productivity at one or more sites as at least 5 months different from EC GPP, or with essentially an opposite seasonal cycle.

Especially the lively models simulate stronger reductions in GPP as the dry season progresses than EC data suggest. Figure 8 summarizes seasonal timing tendencies for models generalized by responsiveness group. In months whose median value of the box plot is above zero, most models at most sites simulate higher GPP than EC estimates. Taller boxes show that the widest spread among models occurs late in the dry season. Mild models on average, shown with the green line in Figure 8, have relatively little systematic difference from EC estimates over the course of a year. Mildness is defined as a dampened seasonal cycle, not by similarity of timing to EC GPP, so this result is not inevitable. Mild models tend to exceed EC estimates slightly in the two months before the driest month, simulating relatively more flourishing early in the dry season than ECs show. In contrast, lively models, whose average the orange line tracks, simulate stronger declines in GPP for 5 months starting with the driest month. For lively models, water stress during the dry season seems more strongly to curtail GPP than EC data suggest is realistic.

At individual sites (Figure S15), overall patterning of relative seasonal timing mostly is similar to the means (Figure S4). The differences between mild and lively models are larger at single sites, with more variation in phase relative to the dry season. On average, lively models differ more from EC estimates than do mild models except at K83. RJA shows little pattern in seasonal differences between models and EC, and mild models match EC GPP timing closely all year. The widespread tendency for models to simulate relative GPP that on average diverges most from EC estimates in the late dry season suggests challenges in modeling soil moisture and/or plant hydraulics.

4. Analysis

4.1. Summary of Findings

Flattening describes the lower variance of predictions than source data when predictive equations are accurately responsive. Either of two mechanisms can cause the variance of simulated outcomes to be nearly as large as the variance of true outcomes despite flattening. At one extreme, the model is accurate in all respects and includes every significant real-world driver. At the other extreme, substantively important drivers are missing and/or their form is misspecified, but variance is inflated by unrealistically strong sensitivity to some or all of the drivers that are included (Text S1). In light of flattening, this study addresses the magnitude of seasonal dynamism in modeled GPP for the Amazonian rainforest, how strongly current weather determines GPP at six EC sites, and the fidelity of modeled seasonal timing to EC seasonality. Lively models are defined as those with higher seasonal amplitude than EC GPP's, while mild models are less seasonal (Figure 4).

Process and statistical models struggle to reproduce EC estimates of Amazonian rainforest GPP. Mean and/or variance of GPP for every model falls outside of 99th percentile confidence intervals (Figure 3). The regression for EC GPP that includes both weather and site-specific intercepts explains a total of 59% of variability (Equation 2), which means that predicted GPP variance from models with accurate sensitivity should be between half and two-thirds as large as EC variance.

Current weather has useful power to explain EC GPP only after site means are considered. With reference data for only six intercepts, this study does not explore the important component of GPP accuracy that resides in site means and their determinants. After allowing for site effects, a linear combination of current month's rain, temperature and light explains only about an eighth (12%) of EC GPP's total variance, equal to

0.38 $gCm^{-2}d^{-1}$. The portion of dynamism explained is low enough to support a qualitative conclusion that the relationship between rainforest GPP and current month's weather is weak.

The weight of evidence suggests that lively models are unrealistically sensitive to weather. Variances for some models that are even higher than EC GPP variance are of special concern. Both simple variance (Figure 3) and seasonal amplitude (Figure 4) for every lively model exceeds its EC equivalent. Flattening implies that instead the model variance and seasonal amplitude should be substantially lower. If modeled variance exceeds its reference equivalent, then model sensitivity to drivers is so excessive that it has overwhelmed flattening. Multiple metrics in this study suggest that lively models are overly responsive, while the mild models are more likely to represent accurately the mean GPP consequences of climate shifts for the Amazon rainforest.

Direct comparisons of descriptive regression slopes are further evidence of excessive sensitivity. Responsiveness to rain is stronger in every lively model than EC rain responsiveness (1.17 average slope vs. 0.48; Figure 6). Perhaps in counterbalance, all statistically significant slopes for temperature and light for each lively model are of the opposite sign from EC GPP's. Finally, for the highly seasonal lively models the annual period of reduced photosynthesis tends to be more severe than for EC GPP and later in the dry season (Figure 8). For lively models, mismatch to EC seasonal cycles is consistent with excessive responsiveness to current rain and perhaps also to cumulative rain stored as soil water. Hypersensitivity means that lively models predict unrealistically large declines in GPP as a consequence of drought.

Whether mild models are overly sensitive to weather is less clear. EC GPP has wider seasonal swings than mild models simulate, and compared to most mild models EC GPP's variance also is larger (Figures 3 and 4). This understated dynamism, or net flattening, makes excessive driver sensitivity less likely. However, it is possible that flattening influences are sufficiently strong to counteract excessive driver responsiveness. Two main contributors to the flattening are likely. One is model misspecification noise due to non-linearity in responses to weather (Figures 7 and S2). Non-linearities were assessed only qualitatively due to MsTMIP's temporal resolution. The second likely cause of low dynamism in mild models is low spread in site intercepts. The range in site means for EC GPP is 4.0 $gCm^{-2}d^{-1}$. Except for one outlier model, the ranges of GPP site means for mild models all are smaller, 0.8 to 3.5. Both model misspecification and low sensitivity to site mean differences could flatten the GPP predictions.

This study's evidence about the accuracy of mild models' responses to current weather is mixed. One test is whether weather predictors as a group explain an appropriate amount of mild model responsiveness. If model sensitivity to weather perfectly matched EC sensitivity, weather would explain the same absolute amount of GPP dynamism as it does for the ECs, 0.38 $gCm^{-2}d^{-1}$. For mild models, the average variance that weather explains is 0.79 $gCm^{-2}d^{-1}$ (range across models = 0.09–4.60). Given the inevitable uncertainty in EC GPP, this check seems at most suggestive that a few mild models respond too weakly but that others may be overresponsive. Direct comparison of driver slopes indicates that the sensitivity of mild models to weather is reasonable overall, although spread among models is considerable. As Figure 6 also shows, average responsiveness to rain and light for mild models is similar to that of EC GPP, while temperature responsiveness is lower but at least of the same sign.

4.2. This Study's Findings Are Generally Consistent With Related Literature

Our results are specific to the scales of time and space for which they are calculated: monthly means for six tower sites between 2000 and 2010. Driver strengths can vary with time integration. At the K67 flux tower, for example, vapor pressure deficit and total and diffuse light largely determined fluctuations in hourly averaged GPP, while a derived metric that also included leaf area index was better at explaining monthly averages (Wu, Guan, et al., 2017). A model that reproduces hourly photosynthetic fluxes well may still have substantial biases at a coarser time scale of annual totals (Keenan et al., 2012). Spatial amalgamation even more strongly reduces apparent variability (Rödig et al., 2018). For example, even though precipitation may drive local dynamism of GPP, gross finitude of global atmospheric water means that temperature largely determines long-term global variability in net land:atmosphere carbon exchange (Jung et al., 2017).

Our analysis agrees with prior findings that rainforest GPP in global vegetation models react too strongly to weather (Ahlström et al., 2017; Baker et al., 2008; Cleveland et al., 2015; Huang et al., 2016; L. Li

et al., 2017; Parazoo et al., 2014; Piao et al., 2013; Poulter et al., 2009; Restrepo-Coupe et al., 2016; von Randow et al., 2013; Z. Zhu et al., 2016). Excessive GPP seasonality was reported for an earlier version of Model J at K67 (Sakaguchi et al., 2011), and for Model I at a flux tower in Guyana (J. Zhu et al., 2018). However, we find also that a few mild models have weak responses to weather.

The mismatch between EC and model responses to temperature is both striking (Figure 6) and important. There is high confidence that tropical temperatures will continue to rise (Jiménez-Muñoz et al., 2013). Disagreement about simulated GPP in months that now are warmer than average represents uncertainty about whether rainforest vegetation will thrive in the near future. Not one model is as welcoming of warmer temperature as is EC GPP. Instead, 10 of the study models have a statistically significant tendency for GPP to decline with increasing temperature, consistent with Huntingford et al. (2013) and Poulter, Aragão, et al. (2010). MacBean et al. (2018) found that the optimal temperature for EBF in one MsTMIP model (Figure 7) is too low. Mismatches in GPP seasonal timing (Figure 8; Text S5; Albert et al., 2018; Albert et al., 2019, Borchert et al., 2002; Doughty & Goulden, 2008; Morton et al., 2016; Samanta et al., 2012; Wilson et al., 2001; Wu et al., 2016), consistent with the findings of Poulter, Hattermann, et al. (2010), suggest that during the dry season, actual plants experience less water stress than is modeled.

Independent indicators of tropical plant response to higher temperatures are limited. Temperature appears to be a positive and stronger driver of net ecosystem exchange globally, and precipitation a weaker driver, than is represented in most dynamic global vegetation models (W. Wang et al., 2013). At La Selva, Costa Rica, the net of GPP and respiration fell strongly with increases in daily minimum temperatures during a 12-year study period (Clark et al., 2013). In contrast, mean annual GPP from 1997 to 2012 showed consistently positive responses of the Normalized Difference Vegetation Index (NDVI) to temperature in the Amazon rainforest (Quetin & Swann, 2017). The authors suggest that positive tropical plant responses to higher temperature may reflect concomitant decreases in cloudiness. However, Moderate Resolution Imaging Spectroradiometer NDVI has accuracy problems in the Amazon (Asner & Alencar, 2010; Hilker et al., 2014). Finally, two other studies are inconclusive about the effect on tropical GPP of high temperatures. Jiménez-Muñoz et al. (2013) establish a correlation between higher temperatures and recent droughts in the Amazon and point to other studies of increased mortality for Amazon trees during droughts, but do not themselves address vegetation responses directly. A CO_2 inversion of net biosphere exchange during 2010–2012 (Alden et al., 2016) that avoids process model priors finds even stronger correlations of net carbon loss with high temperature than with low precipitation for the central Amazon basin. The correlations hold only in the wet season, however, and comparison with two GPP proxies suggest the main determinant may be increased respiration in the heat rather than a GPP response. A challenge in specifying EBF temperature sensitivity is that responses may vary among the thousands of rainforest tree species.

4.3. A Water- Versus Light-Limitation Dichotomy Poorly Characterizes the Models

GPP increases with light in mild models and falls with light in lively models (Figure 6). The lively models also respond more strongly to light. It is tempting to associate each group with either strong response to light at the expense of temperature sensitivity or vice versa.

A trade-off when clouds provide more rain but less light for photosynthesis (Huete et al., 2006; Nemani et al., 2003) has been labeled as light-limited versus water-limited (Arias et al., 2011; Baker et al., 2013, 2019; Myneni et al., 2007). If a place is water-limited, then during sunny periods GPP falls due to greater water stress. If a place is light-limited, then during sunny periods GPP rises, enabled by plentiful water (Graham et al., 2003). The photosynthetic advantage of diffuse compared to direct light direction, however, could dampen potential light limitation due to clouds (Butt et al., 2009). Some observational evidence contradicts the hypothesis that the trade-off is common in the tropics, and instead that GPP may respond minimally to differences in light strength (Restrepo-Coupe et al., 2013).

For the dichotomy to describe a real trade-off for rainforest vegetation, light and rain must be anticorrelated. Increasing tropical cloudiness would reduce light and might increase rain. But in the MsTMIP weather data, rain explains only 3% of the variation in light, and the correlation of EC GPP and light is negative (−0.14), rather than positive as the theory suggests. Neither in the MsTMIP driver data does low rainfall bring more sensible heat in presumed Bowen ratio response to drought stress; rain explains similarly little

(4%) of the variation in temperature. In the lively models GPP does increase strongly with more rain, as would be expected for approaches that focus mostly on water limitations. For four of the nine mild models, GPP is lowest during a dark month for at least one site, but for at least one other site GPP is lowest during a dry month when light is likely to be stronger (Figure S3). The mild models do not appear to be strongly light-limited.

4.4. Weak Response of Modeled Historic GPP to CO_2 Do Not Reveal Future Responses

CO_2 was a significant predictor of GPP only for two especially mild process models (Figure S12). The CO_2 slope for five of the models was statistically indistinguishable from zero. A possible reason is that MIP driver CO_2 and values measured on site are essentially unrelated, with a correlation of −0.11. While this suggests that a descriptive regression of GPP as modeled for an entire cell from MIP driver data is not representative of the six ECs, neither are estimates of CO_2 from the global data set.

Much more than for the other weather drivers, the effect of CO_2 is likely to differ in coming decades from either real or modeled responses in 2000–2010. Over time, changes in GPP due to increasing CO_2 can overwhelm changes due to other environmental drivers because the trend in CO_2 concentration is so persistent and the relative change eventually so large (J. B. Fisher et al., 2013). Also, GPP responses to CO_2 are unlikely to be linear, mainly because multiple and sometimes conflicting components shape a net trade-off between CO_2 fertilization and increased water use efficiency (Clark et al., 2013; Swann et al., 2016). Land surface models differ substantially in how strongly GPP responds to atmospheric CO_2 increases (Piao et al., 2013). As an example, Model D in this study simulates logarithmic increase (Jain et al., 1994).

Other methods than comparison with this EC data set will be needed to assess the accuracy of model sensitivity to CO_2 in rainforest vegetation. Controlled experiments would be optimal, but funding for the first free air carbon enrichment experiment (FACE) in a rainforest ended before towers were built (Amigo, 2020). The connection between GPP and biomass accumulation is complex (Malhi et al., 2015), making the Rainfor plot mensuration network poorly suited to direct studies of photosynthesis rates. However, Rainfor data do reveal shifts in species traits that are consistent with both CO_2 fertilization and drought adaptations despite the sometimes opposite adjustments expected for each (Esquivel-Muelbert et al., 2019). Finally, existing EC data at finer temporal resolution may provide sufficient sample sizes and perhaps wider dynamic ranges to better characterize rainforest GPP sensitivity to both weather drivers and CO_2 concentration.

4.5. Missing Dynamic Predictors

We speculate that important GPP determinants not included in the descriptive regressions include phosphorus availability, leaf demography, soil moisture, and non-linearity of true plant responses to weather. Long-term site productivity differences describe 49% of the total observed variance for ECs. Current weather explains another 12%. The descriptive regressions do not explain the remaining 41% of EC GPP variance that is fluid within a decade. Some of the unexplained variance is response to drivers included in many of the models but not in our descriptive regressions. Examples are leaf area index and carbon stocks in plant tissues, for which the reference data does not allow direct tests. However, very low correlations of model to EC GPP indicate that even if the models do include most of the important drivers, much room for closer simulation remains.

Some processes that field data indicate do influence tropical GPP are included in only a few ESMs. Phosphorus limitation was mentioned earlier. Another is leaf demography and variable degree of seasonal deciduousness. Tropical leaf senescence is heterogeneously responsive to the severity as well as the seasonality of water stress (Albert et al., 2018; Smith et al., 2019; Wu et al., 2018). GPP models are progressing toward responsiveness of tropical photosynthetic capacity to leaf exchange (Caldararu et al., 2012; Manoli et al., 2018; Wu, Serbin, et al., 2017). But the underlying range of natural variability is daunting.

Soil moisture, a lagged indirect consequence (Bloom et al., 2020) of precipitation and temperature, may critically limit GPP even in the wet tropics. One modeling complexity is that root depth distribution and soil water presence each have potential to be the more limiting (Baker et al., 2008). Also, upper basin runoff that accumulates at scales larger than single model cells may cause flooding or change water table depth.

Unfortunately, MsTMIP files do not support direct examination of modeled soil moisture (Text S2). But two indicators suggest that lively models may underestimate root access to deep soil moisture. First, declines in GPP earlier in the dry season in lively models than at ECs imply that real plants tolerate seasonal drought better than lively models simulate. Second, cumulative rainfall, a potential correlate with soil moisture, is a strong GPP predictor at every site. The modeling difficulty is that no lag duration works even reasonably well across all the six tower sites (Text S6; Bonal et al., 2016; Broedel et al., 2017; Corlett, 2016; Feldpausch et al., 2016). The reason is that sites' differing optimal lag periods cancel each other when generalized and blur the importance for GPP of seasonal drought.

It might be possible to test our speculation that lively models underestimate typical root EBF access to deep soil moisture by looking at model simulations for unusual sites. Specifically, a model which limits plant access to deep water storage more tightly than is true at average sites should be most accurate at any sites where there are atypical real limits on plant access to deep soil water. Three of the EC sites qualify. Bedrock is only 2–4 m deep at RJA (von Randow et al., 2004). At CAX, a hardpan layer is 3–5 m deep (Carswell et al., 2002). Flooding makes soil at BAN seasonally anoxic (Christoffersen et al., 2014), which might limit deeper roots' metabolic capability in the wet season. The process models do not have any information about these site particulars that restrict plant access to deep soil water. RJA and BAN are the two sites whose EC GPP the models do simulate most closely. But CAX is on the opposite end of the spectrum, so poorly matched that the average model correlation is negative (Figure S7).

Nearly universal non-linear responses to weather in the models (Figures 7 and S2) imply a solid consensus that GPP's true sensitivity is non-linear. Field studies (Mau et al., 2018; Pau et al., 2018) and theory (Alster et al., 2020; Corlett, 2011) similarly support a premise of non-linearity for at least temperature. Tests of how well the models match the shape of EC responses for tropical forests would therefore be of strong interest. Unfortunately, the EC data used in this study lack enough depth at the temporal resolution of the MsTMIP output to benchmark alternative forms. Our linear descriptive model does justify conclusions that the average responsiveness of many models differs from EC responsiveness, and that compared to ECs, some models simulate either excessive or understated GPP reactions to recent weather. But a simple linear model cannot reveal whether simulations have equal fidelity at very high or very low values of each weather element.

4.6. Fluxcom GPP's Low Variance Illustrates the Analytic Value of Flattening

Statistical models of GPP derive from a limited set of core time series data: satellites, flux towers, and ground-based weather observations. The statistical models included in this study incorporate all three. No fourth-independent data set exists to benchmark the statistical models. In the following example application of the flattening concept to analysis of individual models, some tentative conclusions about each statistical model's accuracy for the Amazon still can be drawn from comparisons to flux towers.

To the best of our knowledge, this study is the first to assess Fluxcom (Model G) specifically for tropical rainforests. One comparison of GPP from 53 EC towers to an early version of Fluxcom included EBF sites but only in Australia and Italy (Joiner et al., 2014). Absence of a Fluxcom assessment that features EBF is conspicuous given that rainforest is the most productive PFT on the planet and for climate perhaps the most important.

The defining source for Fluxcom is EC data. Reassuringly, Fluxcom's gross fit with EC GPP is among the closest ($r = 0.37$, Figure S7). Only Model B has a higher overall correlation with EC GPP. While Fluxcom's mean GPP in site cells is, like all but one model, outside EC credible bounds, it is among the half-dozen models closest to the EC mean. Fluxcom's rain responsiveness slope also is one of the closest to EC estimates (Figure 6). Its scaled temperature slope, 0.006, is much shallower than the slope for EC GPP of 0.460. But so to some degree is every other model's. The sign of Fluxcom's slope for light matches that of the EC towers although its response is stronger by almost half. The month of peak GPP for Fluxcom is within two months of EC estimates for all sites except RJA (Figure S4), again among the best matching of models.

Fluxcom's complex algorithms resemble linear regression in ways that make flattening applicable. For example, in model tree ensembles, a Fluxcom option, machine learning stratifies spatially and temporally defined outcomes into bins. Simulated values for each bin are predictions of a regression based on the bin's members. Fluxcom's weak dynamism globally appears to be an inherent consequence of omitted drivers

and flattening. In our study, Fluxcom's GPP variance (1.4 vs. ECs' 3.3 $gCm^{-2}d^{-1}$) and seasonal amplitude for site cells (2.2 vs. 3.4 $gCm^{-2}d^{-1}$) are about half as wide as tower values. Flattening could help explain Piao et al.'s (2013) finding that Fluxcom GPP is less dynamic than any of 10 DGVMs. Fluxcom and WeCann are less accurate for EBFs than their global average (Alemohammad et al., 2017; Badgley et al., 2019). Reasons for the weaker performance include the dearth of both EC towers and of clear satellite retrievals in the tropics (Jung et al., 2020; Ryu et al., 2019; Tramontana et al., 2016). Fluxcom would thus likely have especially flattened predictions for the tropics.

Fluxcom's globally low GPP variability has been called "undersampled" (Piao et al., 2013), "poorly captured" (Tramontana et al., 2016), underestimated for reasons that are not fully clear (Jung et al., 2020), and, on the product website, "too small" (https://www.bgc-jena.mpg.de/geodb/projects/Data.php). The tendency bears consideration when using Fluxcom. But in light of flattening, we disagree that Fluxcom's mildness is necessarily a weakness. Instead, the low variability suggests theoretically good potential for accurate driver responsiveness.

WeCann has similar strengths and weaknesses as Fluxcom with respect to sensitivity to current weather. The temporal span of WeCann (Model F) prevents reasonable direct comparisons to the EC towers used in this study. But in each comparison across the Amazon basin, WeCann closely resembles Fluxcom (Text S3 and Figures 7, S2, S11–S13). GPP differs at the warmest decile (Figure 7). WeCann is slightly less sensitive to light than is Fluxcom and more so to CO_2 (Figure S12). Current weather plus site intercepts explain less of GPP variance, 80% for Fluxcom and 47% for WeCann (Figure S14). Residual standard errors (Figure S13), which tend to indicate the extent of phase mismatch in a cycle (Taylor, 2001), are similar. The detailed nature of these differences reinforces the overall responsiveness similarity of WeCann and Fluxcom.

The third statistical model, VPM (Model I) is vulnerable to cloud problems for satellite data that are most troublesome in the tropics. While Fluxcom and WeCann feature EC data, VPM emphasizes satellite sources. Gap-filled or missing data overlap heavily with periods of high greenness and potentially of peak rainforest GPP. VPM's unusual spatial pattern of rain responsiveness in the Amazon (Figure S14) corresponds to logical wet season peaks in cloud contamination of satellite data retrievals. Photosynthetically active radiation from the National Centers for Environmental Prediction II weather reanalysis is the multiplicative model component of VPM that cloudiness is most likely to skew. Radiation from weather reanalyzes was specifically omitted as an input to another statistical model due to the product's high uncertainty (Gentine et al., 2019; Jung et al., 2011).

VPM fared poorly in this study's assessments for the Amazon rainforest. VPM is more strongly anticorrelated with EC GPP than any other model ($r = -0.19$). It is the only model for which more rain is associated with lower GPP overall (Figure 6), and in most precipitation deciles (Figure S2). VPM's response to light also is an outlier, increasing at every decile with no saturation (Figure S2). The month with lowest GPP for other models differs from ECs on average by 1.5–3.3 months. VPM averages 4.5 months (Figure S3); The phase of VPM's seasonal cycle is nearly opposite that of EC GPP at most sites. At no site does VPM simulate minimum GPP during a dry season month. While VPM GPP estimates are unrepresentative of the best reference data available for the Amazon (Zhang et al., 2017), a logically underlying reason of cloud contamination applies most strongly to the tropics and could have little effect on VPM's accuracy elsewhere.

Summarizing performance of the statistical models, Fluxcom matches EC GPP only weakly with a correlation of 0.37. But the fit is better than for almost all process models. WeCann compares similarly. The fidelity does not establish the veracity of Fluxcom or WeCann but does verify their anticipated conformity with EC GPP. Among the models assessed, Fluxcom and WeCann appear to be the best wall-to-wall estimates of recent GPP in the Amazon.

4.7. Awareness can Compensate for Some but Not all Problems That Flattening in ESMs Creates

The most important conclusion of this study is that flattening is likely to affect nearly all ESM outputs. Flattening does not imply criticism or limited value of models. Provided responsiveness to drivers is approximately accurate, flattening only describes a cost of the simplification that is the power behind modeling. This section comments on direct consequences of flattened outputs and on benchmarking of variance. The final focus is the special problems that flattening creates for modeled feedbacks.

We emphasize accommodation rather than a goal of eliminating flattening in ESMs. On one hand, the incredibly difficult process of improving ESMs has over time reduced flattening somewhat. Processes and drivers added to the ESMs in Intergovernmental Panel on Climate Change Comprehensive Assessment Reports have not greatly reduced uncertainty around mean temperature trends, but greater complexity has increased the accuracy of simulations (Dahan, 2010). Unless and until models of such large scope as ESMs include accurate equations and data for nearly all process contributors, however, some flattening will remain. Even if such a remarkable level of model complexity could be achieved, completeness might not exceed the value of model simplicity (Levins, 1966). Specific to rainforest GPP, the subtle determinants of seasonality near the equator and difficulty of collecting data in situ mean that progress in modeling probably will remain slower than for other biomes. All of these reasons mean that the important consequences of flattening will persist. Weather sensitivity describes, by definition, the consequences of a changing climate. In evaluating models related to climate, we contend that it is usually better to focus more on sensitivity accuracy than on simulating realistic dynamism.

The simple criterion of modeled variance exceeding benchmark variance is sufficient to establish model hypersensitivity to drivers. Quantifying the exaggeration requires multiplying the benchmarked outcome's variance by the explanatory power of a fully descriptive regression. As this study shows, however, even when uncertainty is significant, qualitatively comparing modeled to benchmarked variance is a useful tool to identify likely hypersensitivity. Accuracy of outcome variances already is integral to International Land Model Benchmarking (ILAMB, Collier et al., 2018) and to many model intercomparison projects (e.g., Houghton et al., 2001; Jupp et al., 2010; F. Li et al., 2019). While ILAMB evaluates responsiveness through relationship metrics, that is, variable-to-variable comparisons independent of time, responsiveness to drivers tends to be even more difficult to benchmark than the variance of outputs.

Workarounds exist to predict future climate variance despite flattening. The options are valuable because climate variability itself may bring risk that warrants analysis (e.g., Bollerslev et al., 1994; Dutta, 2017; Tan, 1998). Consequences of outlier weather include wildland fires, droughts, heat waves, floods, and tropical cyclones (Katz & Brown, 1992). Studies of future extreme weather may predict from absolute change or z-score change in drivers (Abatzoglou & Williams, 2016, e.g.). A conceptually similar option is to build predictions from simulated history (Camargo, 2013). The need to predict outlier values separately is a compliment to the parent model because it reflects potentially accurate driver responsiveness.

The most problematic consequence of flattening is that low dynamism of intermediate model calculations can change predicted mean outcomes. Precipitation intensity in the Amazon rainforest is an illustration. In typical ESM runs, rainfall depth is spread uniformly across a grid cell, for rainforests yielding overabundant mist with few cloudbursts (Baker et al., 2019), or too little dynamism. So much rain then evaporates from tropical leaves that soil water recharge is underestimated. An alternative of cloud superparameterization that distributes rain non-uniformly within land grid cells improves the realism of tropical GPP but may distort modeled precipitation elsewhere (Phillips, 2019). Fire modules in ESMs seem likely to face comparable challenges, responsive as they appropriately are to drought. Flattening potentially affects any modeled feedback loop.

Feasible workarounds to internal distortion from flattening may be scarce. Finer temporal or spatial resolution in effect adds the missing driver of finer-scale heterogeneity. ESM resolution has increased but will remain limited by computational capacity and perhaps the resolution of source data. Fully stochastic modeling has major computational and other complications, and at present is unrealistic for complex climate simulations. Most other solutions add pure or proxy randomness to a subset of drivers, as ensembles do (Farmer & Vogel, 2016). Fuzzy parameters usually manage uncertainty in parameter estimates (Ersoy & Yünsel, 2006; Hoffman & Miller, 1983) but also can be used as a proxy for statistical noise. Stochastic draws from exogenous distributions can be added explicitly (Khodaparast et al., 2008; Pelletier, 1997) to increase model dynamism. Deterministic draws, however, such as from a distribution at calculated percentiles, still exaggerate effective responsiveness. Rare examples of partial stochasticity embedded within full ESMs are cloud superparameterization (Randall, 2013) and the Lund-Potsdam-Jena managed Land model's distribution of monthly rainfall to individual days (Poulter, Aragão, et al., 2010). We do not have more viable solutions for internal flattening but point out the challenge as a valuable area for future research.

5. Conclusions

This study compared 15 process models and three statistical models to GPP estimates from six EC towers in the Amazon rainforest. No model replicated the large differences in mean EC GPP across sites. Ten models had a weaker overall response than EC data to current month's rain, temperature, and light, while eight models had a stronger response to monthly weather than did EC data. Similarity to Amazon flux towers is one of many important ESM accuracy metrics. Assuming that EC GPP is somewhat realistic or better, GPP in lively models is overly responsive to precipitation and in most cases also changes in the opposite direction as the ECs in response to change in light and temperature. Since temperature will continue to rise and rain is likely to become more variable, the liveliest models may substantially exaggerate the Amazon's future change and peril. Nearly accurate deterministic simulation of both sensitivity to drivers and dynamism for Amazonian GPP is unattainable, because current weather explains so little of EC GPP's dynamism. Flattening is therefore a strong influence.

The role of omitted processes and other contributors to error terms in reducing the dynamism of model predictions has wide relevance. The variability-sensitivity trade-off recommends skepticism about model responsiveness when simulated variance is too high. In contrast, low variance of predictions relative to a benchmark shows potential for accurate driver sensitivity and may deserve acclaim.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

GPP for MsTMIP (Huntzinger et al., 2014) was downloaded from https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1225, and driver data (Wei et al., 2014) from https://daac.ornl.gov/NACP/guides/NACP_MsTMIP_Model_Driver.html. For statistical models' GPP, Fluxcom (Jung et al., 2019; Tramontana et al., 2016) was downloaded from https://www.bgc-jena.mpg.de/geodb/projects/Data.php, WeCann (Alemohammad et al., 2017) from https://avdc.gsfc.nasa.gov/pub/data/project/WECANN/, and VPM (Xiao et al., 2005; Zhang et al., 2017) from https://figshare.com/articles/Monthly_GPP_at_0_5_degree/5048011. EC GPP (Restrepo-Coupe et al., 2013) is archived at https://doi.org/10.3334/ORNLDAAC/1842. SiB4 GPP derives from coauthors' prior work.

Acknowledgments
The authors are grateful for the many people who have managed and operated flux towers in the Amazon, developed the models assessed, and participated in MsTMIP. Ideas shared freely at Cross Validated (Stackexchange) helped with myriad details of this study. The authors thank Forrest Hoffman, Mike Ryan, Editor Eleanor Blyth, and two anonymous reviewers for comments that substantially improved the manuscript. EC data assimilation and analysis was funded by the National Aeronautics and Space Administration (NASA) LBA investigation CD-32, and NASA LBA-DMIP project (NNX09AL52G). The authors thank the Department of Energy (DE-SC0014438) and the National Aeronautics and Space Administration (NNX14AI52G) for funding and support.

References

Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, *113*(42), 11770–11775. https://doi.org/10.1073/pnas.1607171113

Ahlström, A., Canadell, J. G., Schurgers, G., Wu, M., Berry, J. A., Guan, K., & Jackson, R. B. (2017). Hydrologic resilience and Amazon productivity. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00306-z

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Albert, L. P., Restrepo-Coupe, N., Smith, M. N., Wu, J., Chavana-Bryant, C., Prohaska, N., et al. (2019). Cryptic phenology in plants: Case studies, implications, and recommendations. *Global Change Biology*, *25*(11), 3591–3608. https://doi.org/10.1111/gcb.14759

Albert, L. P., Wu, J., Prohaska, N., de Camargo, P. B., Huxman, T. E., Tribuzy, E. S., et al. (2018). Age-dependent leaf physiology and consequences for crown-scale carbon uptake during the dry season in an Amazon evergreen forest. *New Phytologist*, *219*(3), 870–884. https://doi.org/10.1111/nph.15056

Alden, C. B., Miller, J. B., Gatti, L. V., Gloor, M. M., Guan, K., Michalak, A. M., et al. (2016). Regional atmospheric CO_2 inversion reveals seasonal and geographic differences in Amazon net biome exchange. *Global Change Biology*, *22*(10), 3427–3443. https://doi.org/10.1111/gcb.13305

Alemohammad, S. H., Fang, B., Konings, A. G., Aires, F., Green, J. K., Kolassa, J., et al. (2017). Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence. *Biogeosciences*, *14*(18), 4101–4124. https://doi.org/10.5194/bg-14-4101-2017

Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., et al. (2010). A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management*, *259*(4), 660–684. https://doi.org/10.1016/j.foreco.2009.09.001

Alster, C. J., von Fischer, J. C., Allison, S. D., & Treseder, K. K. (2020). Embracing a new paradigm for temperature sensitivity of soil microbes. *Global Change Biology*, *26*(6), 3221–3229. https://doi.org/10.1111/gcb.15053

Amigo, I. (2020). The Amazon's fragile future. *Nature*, *578*, 505–507. https://doi.org/10.1038/d41586-020-00508-4

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., et al. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics*, *53*(3), 785–818. https://doi.org/10.1002/2015RG000483

Andreae, M. O. (2002). Biogeochemical cycling of carbon, water, energy, trace gases, and aerosols in Amazonia: The LBA-EUSTACH experiments. *Journal of Geophysical Research*, *107*(D20), 8066. https://doi.org/10.1029/2001JD000524

Aragão, L. E. O. C., Malhi, Y., Roman-Cuesta, R. M., Saatchi, S., Anderson, L. O., & Shimabukuro, Y. E. (2007). Spatial patterns and fire response of recent Amazonian droughts. *Geophysical Research Letters*, *34*(7). https://doi.org/10.1029/2006GL028946

Aragão, L. E. O. C., Poulter, B., Barlow, J. B., Anderson, L. O., Malhi, Y., Saatchi, S., et al. (2014). Environmental change and the carbon balance of Amazonian forests: Environmental change in Amazonia. *Biological Reviews*, *89*(4), 913–931. https://doi.org/10.1111/brv.12088

Arias, P. A., Fu, R., Hoyos, C. D., Li, W., & Zhou, L. (2011). Changes in cloudiness over the Amazon rainforests during the last two decades: Diagnostic and potential causes. *Climate Dynamics*, *37*(5–6), 1151–1164. https://doi.org/10.1007/s00382-010-0903-2

Asner, G. P., & Alencar, A. (2010). Drought impacts on the Amazon forest: The remote sensing perspective. *New Phytologist*, *187*(3), 569–578. https://doi.org/10.1111/j.1469-8137.2010.03310.x

Badgley, G., Anderegg, L. D. L., Berry, J. A., & Field, C. B. (2019). Terrestrial gross primary production: Using NIRV to scale from site to globe. *Global Change Biology*, *25*(11), 3731–3740. https://doi.org/10.1111/gcb.14729

Baker, I. T., Denning, A., Dazlich, D. A., Harper, A. B., Branson, M. D., Randall, D. A., et al. (2019). Surface-atmosphere coupling scale, the fate of water, and ecophysiological function in a Brazilian forest. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2523–2546. https://doi.org/10.1029/2019MS001650

Baker, I. T., Harper, A., da Rocha, H., Denning, A., Araújo, A., Borma, L., et al. (2013). Surface ecophysiological behavior across vegetation and moisture gradients in tropical South America. *Agricultural and Forest Meteorology*, *182–183*, 177–188. https://doi.org/10.1016/j.agrformet.2012.11.015

Baker, I. T., Prihodko, L., Denning, A. S., Goulden, M., Miller, S., & da Rocha, H. R. (2008). Seasonal drought stress in the Amazon: Reconciling models and observations. *Journal of Geophysical Research*, *113*, 1–10. https://doi.org/10.1029/2007JG000644

Barkhordarian, A., von Storch, H., Zorita, E., Loikith, P. C., & Mechoso, C. R. (2017). Observed warming over northern South America has an anthropogenic origin. *Climate Dynamics*. https://doi.org/10.1007/s00382-017-3988-z

Bathiany, S., Dakos, V., Scheffer, M., & Lenton, T. M. (2018). Climate models predict increasing temperature variability in poor countries. *Science Advances*, *4*(5), eaar5809. https://doi.org/10.1126/sciadv.aar5809

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., et al. (2010). Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science*, *329*(5993), 834–838. https://doi.org/10.1126/science.1184984

Bloom, A. A., Bowman, K. W., Liu, J., Konings, A. G., Worden, J. R., Parazoo, N. C., et al. (2020). Lagged effects regulate the inter-annual variability of the tropical carbon balance. *Biogeosciences*, *17*(24), 6393–6422. https://doi.org/10.5194/bg-17-6393-2020

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). Chapter 49: *Arch Models in Handbook of Econometrics* (Vol. 4, pp. 2959–3038). Elsevier. https://doi.org/10.1016/S1573-4412(05)80018-2

Bonal, D., Burban, B., Stahl, C., Wagner, F., & Hérault, B. (2016). The response of tropical rainforests to drought—Lessons from recent research and future prospects. *Annals of Forest Science*, *73*(1), 27–44. https://doi.org/10.1007/s13595-015-0522-5

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., et al. (2011). Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research*, *116*(G2), G02014. https://doi.org/10.1029/2010JG001593

Borchert, R., Rivera, G., & Hagnauer, W. (2002). Modification of vegetative phenology in a tropical semi-deciduous forest by abnormal drought and rain. *Biotropica*, *34*(1), 27–39. https://doi.org/10.1111/j.1744-7429.2002.tb00239.x

Brienen, R. J. W., Phillips, O. L., Feldpausch, T. R., Gloor, E., Baker, T. R., Lloyd, J., et al. (2015). Long-term decline of the Amazon carbon sink. *Nature*, *519*(7543), 344–348. https://doi.org/10.1038/nature14283

Brienen, R. J. W., Schöngart, J., & Zuidema, P. A. (2016). Tree rings in the tropics: Insights into the ecology and climate sensitivity of tropical trees. In G. Goldstein, & L. S. Santiago (Eds.), *Tropical tree physiology: Adaptations and responses in a changing environment* (pp. 439–461). Springer International Publishing. https://doi.org/10.1007/978-3-319-27422-5_20

Broedel, E., Tomasella, J., Cândido, L. A., & von Randow, C. (2017). Deep soil water dynamics in an undisturbed primary forest in central Amazonia: Differences between normal years and the 2005 drought. *Hydrological Processes*, *31*(9), 1749–1759. https://doi.org/10.1002/hyp.11143

Butt, N., New, M., Lizcano, G., & Malhi, Y. (2009). Spatial patterns and recent trends in cloud fraction and cloud-related diffuse radiation in Amazonia. *Journal of Geophysical Research*, *114*(D21). https://doi.org/10.1029/2009JD012217

Caldararu, S., Palmer, P. I., & Purves, D. W. (2012). Inferring Amazon leaf demography from satellite observations of leaf area index. *Biogeosciences*, *9*(4), 1389–1404. https://doi.org/10.5194/bg-9-1389-2012

Camargo, S. J. (2013). Global and regional aspects of tropical cyclone activity in the CMIP5 models. *Journal of Climate*, *26*(24), 9880–9902. https://doi.org/10.1175/JCLI-D-12-00549.1

Carswell, F. E., Costa, A. L., Palheta, M., Malhi, Y., Meir, P., de P. R. Costa, J., et al. (2002). Seasonality in CO_2 and H_2O flux at an eastern Amazonian rain forest. *Journal of Geophysical Research*, *107*(D20), LBA 43-1–LBA 43-16. https://doi.org/10.1029/2000JD000284

Castanho, A. D. A., Coe, M. T., Costa, M. H., Malhi, Y., Galbraith, D., & Quesada, C. A. (2013). Improving simulated Amazon forest biomass and productivity by including spatial variation in biophysical parameters. *Biogeosciences*, *10*(4), 2255–2272. https://doi.org/10.5194/bg-10-2255-2013

Cavaleri, M. A., Reed, S. C., Smith, W. K., & Wood, T. E. (2015). Urgent need for warming experiments in tropical forests. *Global Change Biology*, *21*(6), 2111–2121. https://doi.org/10.1111/gcb.12860

Chadwick, R., Good, P., Martin, G., & Rowell, D. P. (2015). Large rainfall changes consistently projected over substantial areas of tropical land. *Nature Climate Change*, *6*, 177–181. https://doi.org/10.1038/nclimate2805

Chapin, F. S., Woodwell, G. M., Randerson, J. T., Rastetter, E. B., Lovett, G. M., Baldocchi, D. D., et al. (2006). Reconciling carbon-cycle concepts, terminology, and methods. *Ecosystems*, *9*(7), 1041–1050. https://doi.org/10.1007/s10021-005-0105-7

Chevan, A., & Sutherland, M. (1991). Hierarchical partitioning. *The American Statistician*, *45*(2), 90. https://doi.org/10.2307/2684366

Christoffersen, B. O., Restrepo-Coupe, N., Arain, M. A., Baker, I. T., Cestaro, B. P., Ciais, P., et al. (2014). Mechanisms of water supply and vegetation demand govern the seasonality and magnitude of evapotranspiration in Amazonia and Cerrado. *Agricultural and Forest Meteorology*, *191*, 33–50. https://doi.org/10.1016/j.agrformet.2014.02.008

Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., et al. (2021). Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites. *Agricultural and Forest Meteorology*, *301–302*, 108350. https://doi.org/10.1016/j.agrformet.2021.108350

Clark, D. A., Asao, S., Fisher, R., Reed, S., Reich, P. B., Ryan, M. G., et al. (2017). Reviews and syntheses: Field data to benchmark the carbon cycle models for tropical forests. *Biogeosciences*, *14*(20), 4663–4690. https://doi.org/10.5194/bg-14-4663-2017

Clark, D. A., & Clark, D. B. (2011). Assessing tropical forests' climatic sensitivities with long-term data. *Biotropica*, *43*(1), 31–40. https://doi.org/10.1111/j.1744-7429.2010.00654.x

Clark, D. A., Clark, D. B., & Oberbauer, S. F. (2013). Field-quantified responses of tropical rainforest aboveground productivity to increasing CO_2 and climatic stress, 1997-2009. *Journal of Geophysical Research: Biogeosciences*, *118*(2), 783–794. https://doi.org/10.1002/jgrg.20067

Cleveland, C. C., Taylor, P., Chadwick, K. D., Dahlin, K., Doughty, C. E., Malhi, Y., et al. (2015). A comparison of plot- based satellite and Earth system model estimates of tropical forest net primary production. *Global Biogeochemical Cycles*, *29*(5), 626–644. https://doi.org/10.1002/2014GB005022

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al. (2018). The International Land Model Benchmarking (IL-AMB) System: Design, theory, and implementation. *Journal of Advances in Modeling Earth Systems*, 1–24. https://doi.org/10.1029/2018MS001354

Corlett, R. T. (2011). Impacts of warming on tropical lowland rainforests. *Trends in Ecology & Evolution*, *26*(11), 606–613. https://doi.org/10.1016/j.tree.2011.06.015

Corlett, R. T. (2016). The impacts of droughts in tropical forests. *Trends in Plant Science*, *21*(7), 584–593. https://doi.org/10.1016/j.tplants.2016.02.003

Costa, M. H., Cohen, W. (2007). *LBA-ECO CD-15 Leaf Area Index Data (LAI-2000), km 67 Tower Site, Santarem: 2003-2004*. Data set available on-line from LBA Data and Information System. National Institute for Space Research (INPE/CPTEC).

Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., & Luke, C. M. (2013). Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, *494*(7437), 341–344. https://doi.org/10.1038/nature11882

da Rocha, H. R., Manzi, A. O., Cabral, O. M., Miller, S. D., Goulden, M. L., Saleska, S. R., et al. (2009). Patterns of water and heat flux across a biome gradient from tropical forest to savanna in Brazil. *Journal of Geophysical Research*, *114*, 1–8. https://doi.org/10.1029/2007JG000640

Dahan, A. (2010). Putting the Earth System in a numerical box? The evolution from climate modeling toward global change. *Studies in History and Philosophy of Modern Physics*, *41*(3), 282–292. https://doi.org/10.1016/j.shpsb.2010.08.002

Dahlin, K. M., Ponte, D. D., Setlock, E., & Nagelkirk, R. (2017). Global patterns of drought deciduous phenology in semi-arid and savanna-type ecosystems. *Ecography*, *40*(2), 314–323. https://doi.org/10.1111/ecog.02443

Davidson, E. A., de Araújo, A. C., Artaxo, P., Balch, J. K., Brown, I. F., Bustamante, M. C. C., et al. (2012). The Amazon basin in transition. *Nature*, *481*(7381), 321–328. https://doi.org/10.1038/nature10717

Davidson, R., & MacKinnon, J. G. (2004). *Econometric theory and methods*. Oxford University Press.

Domingues, T. F., Berry, J. A., Martinelli, L. A., Ometto, J. P. H. B., & Ehleringer, J. R. (2005). Parameterization of canopy structure and leaf-level gas exchange for an Eastern Amazonian Tropical Rain Forest (Tapajós National Forest, Pará, Brazil). *Earth Interactions*, *9*(17), 1–23. https://doi.org/10.1175/EI149.1

Doughty, C. E., & Goulden, M. L. (2008). Seasonal patterns of tropical forest leaf area index and CO_2 exchange. *Journal of Geophysical Research*, *113*. https://doi.org/10.1029/2007JG000590

Doughty, C. E., Metcalfe, D. B., Girardin, C. A. J., Amézquita, F. F., Cabrera, D. G., Huasco, W. H., et al. (2015). Drought impact on forest carbon dynamics and fluxes in Amazonia. *Nature*, *519*(7541), 78–82. https://doi.org/10.1038/nature14213

Dutta, P. (2017). Modeling of variability and uncertainty in human health risk assessment. *MethodsX*, *4*, 76–85. https://doi.org/10.1016/j.mex.2017.01.005

Ersoy, A., & Yünsel, T. Y. (2006). Geostatistical conditional simulation for the assessment of the quality characteristics of çayırhan lignite deposits. *Energy Exploration & Exploitation*, *24*(6), 391–416. https://doi.org/10.1260/014459806780796312

Esquivel-Muelbert, A., Baker, T. R., Dexter, K. G., Lewis, S. L., Brienen, R. J. W., Feldpausch, T. R., et al. (2019). Compositional response of Amazon forests to climate change. *Global Change Biology*, *25*(1), 39–56. https://doi.org/10.1111/gcb.14413

Farmer, W. H., & Vogel, R. M. (2016). On the deterministic and stochastic use of hydrologic models. *Water Resources Research*, *52*(7), 5619–5633. https://doi.org/10.1002/2016WR019129

Feldpausch, T. R., Phillips, O. L., Brienen, R. J. W., Gloor, E., Lloyd, J., Lopez-Gonzalez, G., et al. (2016). Amazon forest response to repeated droughts. *Global Biogeochemical Cycles*, *30*(7), 964–982. https://doi.org/10.1002/2015GB005133

Feng, X., Porporato, A., & Rodriguez-Iturbe, I. (2013). Changes in rainfall seasonality in the tropics. *Nature Climate Change*, *3*(9), 811–815. https://doi.org/10.1038/nclimate1907

Fisher, J. B., Huntzinger, D. N., Schwalm, C. R., & Sitch, S. (2014). Modeling the terrestrial biosphere. *Annual Review of Environment and Resources*, *39*(1), 91–123. https://doi.org/10.1146/annurev-environ-012913-093456

Fisher, J. B., Sikka, M., Sitch, S., Ciais, P., Poulter, B., Galbraith, D., et al. (2013). African tropical rainforest net carbon dioxide fluxes in the twentieth century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1625), 20120376. https://doi.org/10.1098/rstb.2012.0376

Fisher, R. A., Williams, M., Lobo do Vale, R., Lola da Costa, A., & Meir, P. (2006). Evidence from Amazonian forests is consistent with isohydric control of leaf water potential. *Plant, Cell and Environment*, *29*, 151–165. https://doi.org/10.1111/j.1365-3040.2005.01407.x

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., et al. (2006). Climate-carbon cycle feedback analysis: Results from the C ^{4} MIP model intercomparison. *Journal of Climate*, *19*(14), 3337–3353. https://doi.org/10.1175/JCLI3800.1

Fu, R., Yin, L., Li, W., Arias, P. A., Dickinson, R. E., Huang, L., et al. (2013). Increased dry-season length over southern Amazonia in recent decades and its implication for future climate projection. *Proceedings of the National Academy of Sciences*, *110*(45), 18110–18115. https://doi.org/10.1073/pnas.1302584110

Gatti, L. V., Basso, L. S., Miller, J. B., Gloor, M., Gatti Domingues, L., Cassol, H. L. G., et al. (2021). Amazonia as a carbon source linked to deforestation and climate change. *Nature*, *595*(7867), 388–393. https://doi.org/10.1038/s41586-021-03629-6

Gatti, L. V., Gloor, M., Miller, J. B., Doughty, C. E., Malhi, Y., Domingues, L. G., et al. (2014). Drought sensitivity of Amazonian carbon balance revealed by atmospheric measurements. *Nature*, *506*(7486), 76–80. https://doi.org/10.1038/nature12957

Gentine, P., Massmann, A., Lintner, B. R., Alemohammad, S. H., Fu, R., Green, J. K., et al. (2019). Land-atmosphere interactions in the tropics—A review. *Hydrology and Earth System Sciences*, *23*(10), 4171–4197. https://doi.org/10.5194/hess-23-4171-2019

Girardin, C. A. J., Malhi, Y., Doughty, C. E., Metcalfe, D. B., Meir, P., del Aguila-Pasquel, J., et al. (2016). Seasonal trends of Amazonian rainforest phenology, net primary productivity, and carbon allocation. *Global Biogeochemical Cycles*, *30*(5), 700–715. https://doi.org/10.1002/2015GB005270

Gloor, M., Barichivich, J., Ziv, G., Brienen, R., Schöngart, J., Peylin, P., et al. (2015). Recent Amazon climate as background for possible ongoing and future changes of Amazon humid forests. *Global Biogeochemical Cycles*, *29*(9), 1384–1399. https://doi.org/10.1002/2014GB005080

Gloor, M., Brienen, R. J. W., Galbraith, D., Feldpausch, T. R., Schöngart, J., Guyot, J.-L., et al. (2013). Intensification of the Amazon hydrological cycle over the last two decades. *Geophysical Research Letters*, *40*(9), 1729–1733. https://doi.org/10.1002/grl.50377

Gloor, M., Gatti, L., Brienen, R., Feldpausch, T. R., Phillips, O. L., Miller, J., et al. (2012). The carbon balance of South America: A review of the status, decadal trends and main determinants. *Biogeosciences*, *9*(12), 5407–5430. https://doi.org/10.5194/bg-9-5407-2012

Goulden, M. L., Miller, S. D., da Rocha, H. R., Menton, M. C., de Freitas, H. C., e Silva Figueira, A. M., & de Sousa, C. A. D. (2004). Diel and seasonal patterns of tropical forest CO_2 exchange. *Ecological Applications*, *14*(sp4), 42–54. https://doi.org/10.1890/02-6008

Goulden, M. L., Munger, J. W., Fan, S.-M., Daube, B. C., & Wofsy, S. C. (1996). Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy. *Global Change Biology*, *2*(3), 169–182. https://doi.org/10.1111/j.1365-2486.1996.tb00070.x

Grace, J., Lloyd, J., Mcintyre, J., Miranda, A., Meir, P., Miranda, H., et al. (1995). Fluxes of carbon dioxide and water vapour over an undisturbed tropical forest in south-west Amazonia. *Global Change Biology*, *1*(1), 1–12. https://doi.org/10.1111/j.1365-2486.1995.tb00001.x

Graham, E. A., Mulkey, S. S., Kitajima, K., Phillips, N. G., & Wright, S. J. (2003). Cloud cover limits net CO_2 uptake and growth of a rainforest tree during tropical rainy seasons. *Proceedings of the National Academy of Sciences*, *100*(2), 572–576. https://doi.org/10.1073/pnas.0133045100

Greene, W. H. (2012). Econometric Analysis (7th ed.). Prentice Hall.

Hamby, D. M. (1994). A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models. *Environmental Monitoring and Assessment*, *32*, 135–154. https://doi.org/10.1007/BF00547132

Hawes, J. E., & Peres, C. A. (2016). Patterns of plant phenology in Amazonian seasonally flooded and unflooded forests. *Biotropica*, *48*(4), 465–475. https://doi.org/10.1111/btp.12315

Haynes, K. D., Baker, I. T., Denning, A. S., Stöckli, R., Schaefer, K., Lokupitiya, E. Y., & Haynes, J. M. (2019). Representing grasslands using dynamic prognostic phenology based on biological growth stages: 1. Implementation in the Simple Biosphere Model (SiB4). *Journal of Advances in Modeling Earth Systems*, 1–26. https://doi.org/10.1029/2018MS001540

Haynes, K. D., Baker, I. T., Denning, A. S., Wolf, S., Wohlfahrt, G., Kiely, G., et al. (2019). Representing grasslands using dynamic prognostic phenology based on biological growth stages: Part 2. Carbon cycling. *Journal of Advances in Modeling Earth Systems*, 1–17. https://doi.org/10.1029/2018MS001541

Hilker, T., Lyapustin, A. I., Tucker, C. J., Hall, F. G., Myneni, R. B., Wang, Y., et al. (2014). Vegetation dynamics and rainfall sensitivity of the Amazon. *Proceedings of the National Academy of Sciences*, *111*(45), 16041–16046. https://doi.org/10.1073/pnas.1404870111

Hoffman, F. O., & Miller, C. W. (1983). Uncertainties in environmental radiological assessment models and their implications. In *Annual Meeting of the National Council of Radiation Protection and Measurements* (pp. 1–57).

Houghton, R. A., Lawrence, K. T., Hackler, J. L., & Brown, S. (2001). The spatial distribution of forest biomass in the Brazilian Amazon: A comparison of estimates. *Global Change Biology*, *7*(7), 731–746. https://doi.org/10.1111/j.1365-2486.2001.00426.x

Huang, Y., Gerber, S., Huang, T., & Lichstein, J. W. (2016). Evaluating the drought response of CMIP5 models using global gross primary productivity, leaf area, precipitation, and soil moisture data. *Global Biogeochemical Cycles*, *30*(12), 1827–1846. https://doi.org/10.1002/2016GB005480

Huete, A. R., Didan, K., Shimabukuro, Y. E., Ratana, P., Saleska, S. R., Hutyra, L. R., et al. (2006). Amazon rainforests green-up with sunlight in dry season. *Geophysical Research Letters*, *33*(6). https://doi.org/10.1029/2005GL025583

Huntingford, C., Zelazowski, P., Galbraith, D., Mercado, L. M., Sitch, S., Fisher, R., et al. (2013). Simulated resilience of tropical rainforests to CO_2-induced climate change. *Nature Geoscience*, *6*(4), 268–273. https://doi.org/10.1038/ngeo1741

Huntzinger, D., Schwalm, C., Wei, Y., Cook, R., Michalak, A., Schaefer, K., et al. (2014). *NACP MsTMIP: Global 0.5-deg terrestrial biosphere model outputs in standard format*. https://doi.org/10.3334/ORNLDAAC/1225

Hutyra, L. R., Munger, J. W., Saleska, S. R., Gottlieb, E., Daube, B. C., Dunn, A. L., et al. (2007). Seasonal controls on the exchange of carbon and water in an Amazonian rain forest. *Journal of Geophysical Research*, *112*(G3). https://doi.org/10.1029/2006JG000365

Iwata, H., Malhi, Y., & von Randow, C. (2005). Gap-filling measurements of carbon dioxide storage in tropical rainforest canopy airspace. *Agricultural and Forest Meteorology*, *132*(3–4), 305–314. https://doi.org/10.1016/j.agrformet.2005.08.005

Jain, A. K., Wuebbles, D. J., & Kheshgi, H. S. (1994). Integrated Science Model for Assessment of Climate Change. In *Annual Meeting and Exhibition of the Air and Waste Management Association* (pp. 1–19).

Jiménez-Muñoz, J. C., Sobrino, J. A., Mattar, C., & Malhi, Y. (2013). Spatial and temporal patterns of the recent warming of the Amazon forest. *Journal of Geophysical Research: Atmospheres*, *118*(11), 5204–5215. https://doi.org/10.1002/jgrd.50456

Joiner, J., Yoshida, Y., Vasilkov, A., Schaefer, K., Jung, M., Guanter, L., et al. (2014). The seasonal cycle of satellite chlorophyll fluorescence observations and its relationship to vegetation phenology and ecosystem atmosphere carbon exchange. *Remote Sensing of Environment*, *152*, 375–391. https://doi.org/10.1016/j.rse.2014.06.022

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., et al. (2019). The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Scientific Data*, *6*(1), 74. https://doi.org/10.1038/s41597-019-0076-8

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, *116*(G3). https://doi.org/10.1029/2010JG001566

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., et al. (2017). Compensatory water effects link yearly global land CO_2 sink changes to temperature. *Nature*, *541*(7638), 516–520. https://doi.org/10.1038/nature20780

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., et al. (2020). Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, *17*(5), 1343–1365. https://doi.org/10.5194/bg-17-1343-2020

Jupp, T. E., Cox, P. M., Rammig, A., Thonicke, K., Lucht, W., & Cramer, W. (2010). Development of probability density functions for future South American rainfall. *New Phytologist*, *187*(3), 682–693. https://doi.org/10.1111/j.1469-8137.2010.03368.x

Katz, R. W., & Brown, B. G. (1992). Extreme Events in a Changing Climate: Variability is More Important than Averages. *Climatic Change*, *21*, 289–302. https://doi.org/10.1007/BF00139728

Keenan, T. F., Davidson, E., Moffat, A. M., Munger, W., & Richardson, A. D. (2012). Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling. *Global Change Biology*, *18*(8), 2555–2569. https://doi.org/10.1111/j.1365-2486.2012.02684.x

Keenan, T. F., Hollinger, D. Y., Bohrer, G., Dragoni, D., Munger, J. W., Schmid, H. P., & Richardson, A. D. (2013). Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature*, *499*(7458), 324–327. https://doi.org/10.1038/nature12291

Khodaparast, H. H., Mottershead, J. E., & Friswell, M. I. (2008). Perturbation methods for the estimation of parameter variability in stochastic model updating. *Mechanical Systems and Signal Processing*, *22*(8), 1751–1773. https://doi.org/10.1016/j.ymssp.2008.03.001

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., et al. (2013). The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences*, *10*(11), 7109–7131. https://doi.org/10.5194/bg-10-7109-2013

Levine, N. M., Zhang, K., Longo, M., Baccini, A., Phillips, O. L., Lewis, S. L., et al. (2016). Ecosystem heterogeneity determines the ecological resilience of the Amazon to climate change. *Proceedings of the National Academy of Sciences*, *113*(3), 793–797. https://doi.org/10.1073/pnas.1511344112

Levins, R. (1966). The Strategy of Model Building in Population Biology. *American Scientist*, *54*(4), 421–431. Retrieved from www.jstor.org/stable/27836590

Li, F., Martin, M. V., Andreae, M. O., Arneth, A., Hantson, S., Kaiser, J. W., et al. (2019). Historical (1700-2012) global multi-model estimates of the fire emissions from the Fire Modeling Intercomparison Project (FireMIP). *Atmospheric Chemistry and Physics*, *19*(19), 12545–12567. https://doi.org/10.5194/acp-19-12545-2019

Li, L., Wang, Y.-P., Beringer, J., Shi, H., Cleverly, J., Cheng, L., et al. (2017). Responses of LAI to rainfall explain contrasting sensitivities to carbon uptake between forest and non-forest ecosystems in Australia. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-11063-w

Li, W., Fu, R., & Dickinson, R. E. (2006). Rainfall and its seasonality over the Amazon in the 21st century as assessed by the coupled models for the IPCC AR4. *Journal of Geophysical Research: Atmospheres*, *111*(D2). https://doi.org/10.1029/2005JD006355

Li, W., Fu, R., Juárez, R. I. N., & Fernandes, K. (2008). Observed change of the standardized precipitation index, its potential cause and implications to future climate change in the Amazon region. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1498), 1767–1772. https://doi.org/10.1098/rstb.2007.0022

Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., et al. (2017). Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño. *Science*, *358*(6360), eaam5690. https://doi.org/10.1126/science.aam5690

Lopes, A. V., Chiang, J. C. H., Thompson, S. A., & Dracup, J. A. (2016). Trend and uncertainty in spatial-temporal patterns of hydrological droughts in the Amazon basin: Hydrological droughts in the Amazon. *Geophysical Research Letters*, *43*(7), 3307–3316. https://doi.org/10.1002/2016GL067738

MacBean, N., Maignan, F., Bacour, C., Lewis, P., Peylin, P., Guanter, L., et al. (2018). Strong constraint on modelled global carbon uptake using solar-induced chlorophyll fluorescence data. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-20024-w

Malavelle, F. F., Haywood, J. M., Mercado, L. M., Folberth, G. A., Bellouin, N., Sitch, S., & Artaxo, P. (2019). Studying the impact of biomass burning aerosol radiative and climate effects on the Amazon rainforest productivity with an Earth system model. *Atmospheric Chemistry and Physics*, *19*(2), 1301–1326. https://doi.org/10.5194/acp-19-1301-2019

Malhi, Y., Aragão, L. E. O. C., Galbraith, D., Huntingford, C., Fisher, R., Zelazowski, P., et al. (2009). Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest. *Proceedings of the National Academy of Sciences*, *106*(49), 20610–20615. https://doi.org/10.1073/pnas.0804619106

Malhi, Y., Aragão, L. E. O. C., Metcalfe, D. B., Paiva, R., Quesada, C. A., Almeida, S. S., et al. (2009). Comprehensive assessment of carbon productivity, allocation and storage in three Amazonian forests. *Global Change Biology*, *15*(5), 1255–1274. https://doi.org/10.1111/j.1365-2486.2008.01780.x

Malhi, Y., Doughty, C. E., Goldsmith, G. R., Metcalfe, D. B., Girardin, C. A. J., Marthews, T. R., et al. (2015). The linkages between photosynthesis, productivity, growth and biomass in lowland Amazonian forests. *Global Change Biology*, *21*(6), 2283–2295. https://doi.org/10.1111/gcb.12859

Malhi, Y., & Wright, J. (2004). Spatial patterns and recent trends in the climate of tropical rainforest regions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1443), 311–329. https://doi.org/10.1098/rstb.2003.1433

Manoli, G., Ivanov, V. Y., & Fatichi, S. (2018). Dry-season greening and water stress in Amazonia: The role of modeling leaf phenology. *Journal of Geophysical Research: Biogeosciences*, *123*(6), 1909–1926. https://doi.org/10.1029/2017JG004282

Mau, A., Reed, S., Wood, T., & Cavaleri, M. (2018). Temperate and tropical forest canopies are already functioning beyond their thermal thresholds for photosynthesis. *Forests*, *9*(1), 47. https://doi.org/10.3390/f9010047

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., et al. (2012). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, *4*(3). https://doi.org/10.1029/2012MS000154

Miller, S. D., Goulden, M. L., Menton, M. C., da Rocha, H. R., de Freitas, H. C., Figueira, A. M. e S., & de Sousa, C. A. D. (2004). Biometric and micrometeorological measurements of tropical forest carbon balance. *Ecological Applications*, *14*(sp4), 114–126. https://doi.org/10.1890/02-6005

Morton, D. C., Rubio, J., Cook, B. D., Gastellu-Etchegorry, J.-P., Longo, M., Choi, H., et al. (2016). Amazon forest structure generates diurnal and seasonal variability in light utilization. *Biogeosciences*, *13*(7), 2195–2206. https://doi.org/10.5194/bg-13-2195-2016

Murray-Tortarolo, G., Friedlingstein, P., Sitch, S., Seneviratne, S. I., Fletcher, I., Mueller, B., et al. (2016). The dry season intensity as a key driver of NPP trends: Dry season and vegetation productivity. *Geophysical Research Letters*, *43*(6), 2632–2639. https://doi.org/10.1002/2016GL068240

Myneni, R. B., Yang, W., Nemani, R. R., Huete, A. R., Dickinson, R. E., Knyazikhin, Y., et al. (2007). Large seasonal swings in leaf area of Amazon rainforests. *Proceedings of the National Academy of Sciences*, *104*(12), 4820–4823. https://doi.org/10.1073/pnas.0611338104

Mystakidis, S., Davin, E. L., Gruber, N., & Seneviratne, S. I. (2016). Constraining future terrestrial carbon cycle projections using observation-based water and carbon flux estimates. *Global Change Biology*, *22*(6), 2198–2215. https://doi.org/10.1111/gcb.13217

Negrón Juárez, R. I., da Rocha, H. R., e Figueira, A. M. S., Goulden, M. L., & Miller, S. D. (2009). An improved estimate of leaf area index based on the histogram analysis of hemispherical photographs. *Agricultural and Forest Meteorology*, *149*(6–7), 920–928. https://doi.org/10.1016/j.agrformet.2008.11.012

Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., et al. (2003). Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science*, *300*(5625), 1560–1563. https://doi.org/10.1126/science.1082750

Oleson, K. W., & Lawrence, D. M. (2013). *Technical description of version 4.5 of the Community Land Model (CLM)*. Technical Note TN-503+STR (p. 434). NCAR. Retrieved from https://www.cesm.ucar.edu/models/cesm1.2/clm/CLM45_Tech_Note.pdf

Oliveira, R. S., Dawson, T. E., Burgess, S. S. O., & Nepstad, D. C. (2005). Hydraulic redistribution in three Amazonian trees. *Oecologia*, *145*(3), 354–363. https://doi.org/10.1007/s00442-005-0108-2

Panisset, J. S., Libonati, R., Gouveia, C. M. P., Machado-Silva, F., França, D. A., França, J. R. A., & Peres, L. F. (2018). Contrasting patterns of the extreme drought episodes of 2005, 2010 and 2015 in the Amazon Basin. *International Journal of Climatology*, *38*(2), 1096–1104. https://doi.org/10.1002/joc.5224

Parazoo, N. C., Bowman, K., Fisher, J. B., Frankenberg, C., Jones, D. B. A., Cescatti, A., et al. (2014). Terrestrial gross primary production inferred from satellite fluorescence and vegetation models. *Global Change Biology*, *20*(10), 3103–3121. https://doi.org/10.1111/gcb.12652

Pau, S., Detto, M., Kim, Y., & Still, C. J. (2018). Tropical forest temperature thresholds for gross primary productivity. *Ecosphere*, *9*(7), e02311. https://doi.org/10.1002/ecs2.2311

Pelletier, J. D. (1997). Analysis and modeling of the natural variability of climate. *Journal of Climate*, *10*, 1331–1342. https://doi.org/10.1175/1520-0442(1997)010<1331:AAMOTN>2.0.CO;2

Phillips, M. C. (2019). *Multiple scales of surface-atmosphere coupling in an earth system model*. Doctoral. Colorado State University.
Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., et al. (2013). Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends. *Global Change Biology*, *19*(7), 2117–2132. https://doi.org/10.1111/gcb.12187
Poulter, B., Aragão, L., Heyder, U., Gumpenberger, M., Heinke, J., Langerwisch, F., et al. (2010). Net biome production of the Amazon Basin in the 21st century. *Global Change Biology*, *16*(7), 2062–2075. https://doi.org/10.1111/j.1365-2486.2009.02064.x
Poulter, B., Hattermann, F., Hawkins, E., Zaehle, S., Sitch, S., Restrepo-Coupe, N., et al. (2010). Robust dynamics of Amazon dieback to climate change with perturbed ecosystem model parameters. *Global Change Biology*, *16*(9), 2476–2495. https://doi.org/10.1111/j.1365-2486.2009.02157.x
Poulter, B., Heyder, U., & Cramer, W. (2009). Modeling the sensitivity of the seasonal cycle of GPP to dynamic LAI and soil depths in tropical rainforests. *Ecosystems*, *12*(4), 517–533. https://doi.org/10.1007/s10021-009-9238-4
Pyle, E. H., Santoni, G. W., Nascimento, H. E. M., Hutyra, L. R., Vieira, S., Curran, D. J., et al. (2008). Dynamics of carbon, biomass, and structure in two Amazonian forests. *Journal of Geophysical Research*, *113*(G1), 1–20. https://doi.org/10.1029/2007JG000592
Quesada, C. A., Lloyd, J., Schwarz, M., Baker, T. R., Phillips, O. L., Patiño, S., et al. (2009). Regional and large-scale patterns in Amazon forest structure and function are mediated by variations in soil physical and chemical properties. *Biogeosciences Discussions*, *6*(2), 3993–4057. https://doi.org/10.5194/bgd-6-3993-2009
Quetin, G. R., & Swann, A. L. S. (2017). Empirically derived sensitivity of vegetation to climate across global gradients of temperature and precipitation. *Journal of Climate*, *30*(15), 5835–5849. https://doi.org/10.1175/JCLI-D-16-0829.1
Rammig, A., Jupp, T., Thonicke, K., Tietjen, B., Heinke, J., Ostberg, S., et al. (2010). Estimating the risk of Amazonian forest dieback. *New Phytologist*, *187*(3), 694–706. https://doi.org/10.1111/j.1469-8137.2010.03318.x
Randall, D. A. (2013). Beyond deadlock. *Geophysical Research Letters*, *40*(22), 5970–5976. https://doi.org/10.1002/2013GL057998
Rap, A., Spracklen, D. V., Mercado, L., Reddington, C. L., Haywood, J. M., Ellis, R. J., et al. (2015). Fires increase Amazon forest productivity through increases in diffuse radiation. *Geophysical Research Letters*, *42*(11), 4654–4662. https://doi.org/10.1002/2015GL063719
Restrepo-Coupe, N., da Rocha, H. R., Hutyra, L. R., de Araujo, A. C., Borma, L. S., Christoffersen, B., et al. (2021). LBA-ECO CD-32 Flux Tower Network Data Compilation, Brazilian Amazon: 1999-2006, V2. Oak Ridge, TN: ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/1842
Restrepo-Coupe, N., Hutyra, L. R., da Araujo, A. C., Borma, L. S., Christoffersen, B., Cabral, O. M., et al. (2013). What drives the seasonality of photosynthesis across the Amazon basin? A cross-site analysis of eddy flux tower measurements from the Brasil flux network. *Agricultural and Forest Meteorology*, *182–183*, 128–144. https://doi.org/10.1016/j.agrformet.2013.04.031
Restrepo-Coupe, N., Levine, N. M., Christoffersen, B. O., Albert, L. P., Wu, J., Costa, M. H., et al. (2016). Do dynamic global vegetation models capture the seasonality of carbon fluxes in the Amazon basin? A data-model intercomparison. *Global Change Biology*, *23*(1), 191–208. https://doi.org/10.1111/gcb.13442
Rice, A. H., Pyle, E. H., Saleska, S. R., Hutyra, L., Palace, M., Keller, M., et al. (2004). Carbon balance and vegetation dynamics in an old-growth Amazonian Forest. *Ecological Applications*, *14*(sp4), 55–71. https://doi.org/10.1890/02-6006
Rödig, E., Cuntz, M., Rammig, A., Fischer, R., Taubert, F., & Huth, A. (2018). The importance of forest structure for carbon fluxes of the Amazon rainforest. *Environmental Research Letters*, *13*(5), 054013. https://doi.org/10.1088/1748-9326/aabc61
Rogers, A., Medlyn, B. E., Dukes, J. S., Bonan, G., von Caemmerer, S., Dietze, M. C., et al. (2017). A roadmap for improving the representation of photosynthesis in Earth system models. *New Phytologist*, *213*(1), 22–42. https://doi.org/10.1111/nph.14283
Ryan, M. G. (2011). Tree responses to drought. *Tree Physiology*, *31*(3), 237–239. https://doi.org/10.1093/treephys/tpr022
Ryu, Y., Berry, J. A., & Baldocchi, D. D. (2019). What is global photosynthesis? History, uncertainties and opportunities. *Remote Sensing of Environment*, *223*, 95–114. https://doi.org/10.1016/j.rse.2019.01.016
Sakaguchi, K., Zeng, X., Christoffersen, B. J., Restrepo-Coupe, N., Saleska, S. R., & Brando, P. M. (2011). Natural and drought scenarios in an east central Amazon forest: Fidelity of the Community Land Model 3.5 with three biogeochemical models. *Journal of Geophysical Research*, *116*(G1). https://doi.org/10.1029/2010JG001477
Saleska, S. R., Miller, S. D., Matross, D. M., Goulden, M. L., Wofsy, S. C., da Rocha, H. R., et al. (2003). Carbon in Amazon forests: Unexpected seasonal fluxes and disturbance-induced losses. *Science*, *302*(5650), 1554–1557. https://doi.org/10.1126/science.1091165
Samanta, A., Knyazikhin, Y., Xu, L., Dickinson, R. E., Fu, R., Costa, M. H., et al. (2012). Seasonal changes in leaf area of Amazon forests from leaf flushing and abscission. *Journal of Geophysical Research*, *117*(G1). https://doi.org/10.1029/2011JG001818
Smith, M. N., Stark, S. C., Taylor, T. C., Ferreira, M. L., Oliveira, E., Restrepo-Coupe, N., et al. (2019). Seasonal and drought-related changes in leaf area profiles depend on height and light environment in an Amazon forest. *New Phytologist*, *222*(3), 1284–1297. https://doi.org/10.1111/nph.15726
Swann, A. L. S., Hoffman, F. M., Koven, C. D., & Randerson, J. T. (2016). Plant responses to increasing CO_2 reduce estimates of climate impacts on drought severity. *Proceedings of the National Academy of Sciences*, *113*(36), 10019–10024. https://doi.org/10.1073/pnas.1604581113
Tan, B. (1998). Agile manufacturing and management of variability. *International Transactions in Operational Research*, *5*(5), 375–388. https://doi.org/10.1111/j.1475-3995.1998.tb00121.x
Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, *106*(D7), 7183–7192. https://doi.org/10.1029/2000JD900719
Thomas, R. Q., Zaehle, S., Templer, P. H., & Goodale, C. L. (2013). Global patterns of nitrogen limitation: Confronting two global biogeochemical models with observations. *Global Change Biology*, *19*(10), 2986–2998. https://doi.org/10.1111/gcb.12281
Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., et al. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, *13*(14), 4291–4313. https://doi.org/10.5194/bg-13-4291-2016
van der Sleen, P., Groenendijk, P., Vlam, M., Anten, N. P. R., Boom, A., Bongers, F., et al. (2015). No growth stimulation of tropical trees by 150 years of CO_2 fertilization but water-use efficiency increased. *Nature Geoscience*, *8*(1), 24–28. https://doi.org/10.1038/ngeo2313
Vicari, A. S., Mokhtari, A., Morales, R. A., Jaykus, L.-A., Frey, H. C., Slenning, B. D., & Cowen, P. (2007). Second-order modeling of variability and uncertainty in microbial hazard characterization. *Journal of Food Protection*, *70*(2), 363–372. https://doi.org/10.4315/0362-028X-70.2.363
Vogel, R. M. (1999). Stochastic and deterministic world views. *Journal of Water Resources Planning and Management*, *125*(6), 311–313. https://doi.org/10.1061/(asce)0733-9496(1999)125:6(311)
von Randow, C., Manzi, A. O., Kruijt, B., de Oliveira, P. J., Zanchi, F. B., Silva, R. L., et al. (2004). Comparative measurements and seasonal variations in energy and carbon exchange over forest and pasture in South West Amazonia. *Theoretical and Applied Climatology*, *78*(1). https://doi.org/10.1007/s00704-004-0041-z

von Randow, C., Zeri, M., Restrepo-Coupe, N., Muza, M. N., de Gonçalves, L. G. G., Costa, M. H., et al. (2013). Inter-annual variability of carbon and water fluxes in Amazonian forest, Cerrado and pasture sites, as simulated by terrestrial biosphere models. *Agricultural and Forest Meteorology*, *182–183*, 145–155. https://doi.org/10.1016/j.agrformet.2013.05.015

Wang, W., Ciais, P., Nemani, R. R., Canadell, J. G., Piao, S., Sitch, S., et al. (2013). Variations in atmospheric CO_2 growth rates coupled with tropical temperature. *Proceedings of the National Academy of Sciences*, *110*(32), 13061–13066. https://doi.org/10.1073/pnas.1219683110

Wang, X., Piao, S., Ciais, P., Friedlingstein, P., Myneni, R. B., Cox, P., et al. (2014). A two-fold increase of carbon cycle sensitivity to tropical temperature variations. *Nature*, *506*(7487), 212–215. https://doi.org/10.1038/nature12915

Wei, Y., Liu, S., Huntzinger, D., Michalak, A., Viovy, N., Post, W., et al. (2014). *NACP MsTMIP: Global and north American driver data for multi-model intercomparison*. https://doi.org/10.3334/ORNLDAAC/1220

Wielicki, B. A., Wong, T., Allan, R. P., Slingo, A., Kiehl, J. T., Soden, B. J., et al. (2002). Evidence for large decadal variability in the tropical mean radiative energy budget. *Science*, *295*(5556), 841–844. https://doi.org/10.1126/science.1065837

Wilson, K. B., Baldocchi, D. D., & Hanson, P. J. (2001). Leaf age affects the seasonal pattern of photosynthetic capacity and net ecosystem exchange of carbon in a deciduous forest. *Plant, Cell and Environment*, *24*(6), 571–583. https://doi.org/10.1046/j.0016-8025.2001.00706.x

Wu, J., Albert, L. P., Lopes, A. P., Restrepo-Coupe, N., Hayek, M., Wiedemann, K. T., et al. (2016). Leaf development and demography explain photosynthetic seasonality in Amazon evergreen forests. *Science*, *351*(6276), 972–976. https://doi.org/10.1126/science.aad5068

Wu, J., Guan, K., Hayek, M., Restrepo-Coupe, N., Wiedemann, K. T., Xu, X., et al. (2017). Partitioning controls on Amazon forest photosynthesis between environmental and biotic factors at hourly to interannual timescales. *Global Change Biology*, *23*(3), 1240–1257. https://doi.org/10.1111/gcb.13509

Wu, J., Kobayashi, H., Stark, S. C., Meng, R., Guan, K., Tran, N. N., et al. (2018). Biological processes dominate seasonality of remotely sensed canopy greenness in an Amazon evergreen forest. *New Phytologist*, *217*(4), 1507–1520. https://doi.org/10.1111/nph.14939

Wu, J., Serbin, S. P., Xu, X., Albert, L. P., Chen, M., Meng, R., et al. (2017). The phenology of leaf quality and its within-canopy variation is essential for accurate modeling of photosynthesis in tropical evergreen forests. *Global Change Biology*, *23*(11), 4814–4827. https://doi.org/10.1111/gcb.13725

Xiao, X., Zhang, Q., Saleska, S., Hutyra, L., De Camargo, P., Wofsy, S., et al. (2005). Satellite-based modeling of gross primary production in a seasonally moist tropical evergreen forest. *Remote Sensing of Environment*, *94*(1), 105–122. https://doi.org/10.1016/j.rse.2004.08.015

Xu, L., Saatchi, S. S., Yang, Y., Myneni, R. B., Frankenberg, C., Chowdhury, D., & Bi, J. (2015). Satellite observation of tropical forest seasonality: Spatial patterns of carbon exchange in Amazonia. *Environmental Research Letters*, *10*(8), 084005. https://doi.org/10.1088/1748-9326/10/8/084005

Yan, H., Wang, S.-Q., da Rocha, H. R., Rap, A., Bonal, D., Butt, N., et al. (2017). Simulation of the unexpected photosynthetic seasonality in Amazonian evergreen forests by using an improved diffuse fraction-based light use efficiency model. *Journal of Geophysical Research: Biogeosciences*, *122*(11), 3014–3030. https://doi.org/10.1002/2017JG004008

Yang, S., Li, Z., Yu, J.-Y., Hu, X., Dong, W., & He, S. (2018). El Niño-Southern Oscillation and its impact in the changing climate. *National Science Review*, *5*(6), 840–857. https://doi.org/10.1093/nsr/nwy046

Yang, X., Thornton, P. E., Ricciuto, D. M., & Post, W. M. (2014). The role of phosphorus dynamics in tropical forests—A modeling study using CLM-CNP. *Biogeosciences*, *11*(6), 1667–1681. https://doi.org/10.5194/bg-11-1667-2014

Zhang, Y., Xiao, X., Jin, C., Dong, J., Zhou, S., Wagle, P., et al. (2016). Consistency between sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America. *Remote Sensing of Environment*, *183*, 154–169. https://doi.org/10.1016/j.rse.2016.05.015

Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., & Dong, J. (2017). A global moderate resolution dataset of gross primary production of vegetation for 2000–2016. *Scientific Data*, *4*, 170165. https://doi.org/10.1038/sdata.2017.165

Zhu, J., Zhang, M., Zhang, Y., Zeng, X., & Xiao, X. (2018). Response of tropical terrestrial gross primary production to the Super El Niño event in 2015. *Journal of Geophysical Research: Biogeosciences*, *123*(10), 3193–3203. https://doi.org/10.1029/2018JG004571

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., et al. (2016). Greening of the Earth and its drivers. *Nature Climate Change*, *6*(8), 791–795. https://doi.org/10.1038/nclimate3004